联邦元学习的稳健推断

Zijian Guo罗格斯大学，皮斯卡塔韦，美国Xiudi Li哈佛大学，波士顿，美国Larry Han哈佛大学，波士顿，美国Tianxi Cai哈佛大学，波士顿，美国

arXiv:2301.00718v1 [stat.ME] 2 Jan 2023

摘要。综合利用多个数据来源的信息，对于确保知识的普遍适用性至关重要。由于不同来源之间存在异质性，且隐私顾虑带来数据共享限制，多源数据的整合分析面临挑战。本文针对来自多个站点的数据，提出一个通用的联邦元学习稳健推断框架，可对主导模型进行统计推断；主导模型定义为与多数站点相匹配的模型。对主导模型进行统计推断颇具挑战，因为这需要借助数据自适应机制选择合格站点，并在后续分析中考虑选择不确定性。我们提出一种新颖的抽样方法，以处理由选择带来的额外变异。我们设计的置信区间构造无需站点共享个体层面数据，并且即使合格站点的选择并非完全准确，也能保证其有效性。所提出的联邦元学习稳健推断（RIFL）方法适用范围广，并通过三个推断问题加以说明：参数模型的聚合、高维预测模型，以及平均处理效应的推断。我们使用 RIFL，基于来自四个国家、代表 275 家医院的 16 个医疗中心的真实世界电子健康记录数据，对 COVID-19 住院患者的死亡风险开展联邦学习。

*Ke*关键词：后选择推断；异质性数据；多源数据；隐私保护 - ing; 高维推断。

1. 引言

众包，即汇集群体智慧以解决问题的过程，是一种有益的社区方法，可提升从教育 [Heffernan和Heffernan， 2014]到公共卫生 [Han等人，2018年，Wang等人， 2020]等领域的决策能力。与由单一群体制定的传统专家主导的解决方案相比，纳入多个多元群体的意见能够提高最终决策的质量 [Surowiecki， 2005]。在健康研究中，众包促成了大流行病期间新药的发现 [Chodera等人， 2020], 、以患者为中心的乳腺摄影报告的设计 [Short等人， 2017], 以及用于对放射治疗中的肿瘤进行分类的机器学习算法的开发 [Mak等人， 2019]。

“群体智慧”这一现象背后蕴含着这样一种统计学与哲学观念：从多个数据来源中学习是有益的。与仅从

单一数据源中学习相比，整合来自不同数据源的信息可以提高研究发现的可推广性和可迁移性。由于数据质量不佳，或底层数据生成过程存在异质性，来自单一数据源的研究发现可能无法推广到所关注的新目标总体。

整合多个数据来源的数据，可能是仅使用单一数据源的一种有价值的替代方案。然而，将多个数据来源直接合并到单个数据集中进行分析，往往效果不佳，甚至不可行。不同数据源之间的异质性可能会严重影响采用此类汇总分析策略所得预测或推断的准确性 [Leek 等，2010，Ling 等，2022]。作为汇总分析的替代方案，荟萃分析常用于综合多项研究的信息。标准荟萃分析方法会汇总多项研究的定量证据。随机效应荟萃分析等荟萃分析变体已被用于探究研究间异质性、发表偏倚等潜在偏差，以及小样本研究效应。然而，大多数现有的、能够考虑异质性的荟萃分析工具都要求采用强建模假设，且不考虑当某些站点的数据分布与其他站点存在显著差异时，推断是否有效。

另一个尤为重要的挑战是生物医学研究中的数据隐私问题。美国的法规（如《健康保险流通与责任法案》（HIPAA）隐私规则）以及欧盟的法规（如《通用数据保护条例》（GDPR）和欧洲药品管理局（EMA）隐私声明）保护患者的个人信息，并禁止在站点之间传输患者级数据。这些法规使实现整合数据分析的愿景更加困难，也凸显了对无需共享个体层面数据的联邦式整合分析方法的需求。

当研究间异质性显著且存在离群值时，整合分析的一种理想策略是识别主导模型，以实现共识学习。主导模型定义为多数站点所满足的模型。识别主导模型可以直观地通过多数规则实现 [Sorkin 等，1998年，Kerr 等，2004年，Hastie 和 Kameda，2005], 该规则选择超过半数个体认同的选项。多数规则广泛应用于现代自由民主政体，也用于各种研究领域。例如，基因组学数据通常被分为多个批次，但批次间的异质性可能导致数据出现不理想的变异 [Leek 等，2010年，Ling 等， 2022]。在这种情形下，目标是识别与多数批次一致性较低的批次，并在后续分析中对这类差异进行调整 [Trippa 等， 2015]。另一个例子是临床试验设计：通常无法或不宜将患者纳入对照组。在这种情况下，可以利用历史试验或观察性研究的数据构建外部对照组 [Jahanshahi 等，2021年，Ventz 等，2019年，Davi 等， 2020]。然而，当存在许多此类历史数据来源时，谨慎选择与其他多数数据来源相似性较高的数据来源至关重要。最后一个例子是孟德尔随机化，其中使用多个遗传标记作为工具变量（IVs），以控制潜在的未测量混杂因素。每个工具变量都有其对应的因果效应估计量，研究目标是识别与多数估计效应相匹配的因果效应 [Burgess 等，2017年，Bowden 等，2016年，Kang 等， 2016]。

在缺乏关于主导模型的先验知识时，采用数据自适应方法来选择合适的站点，以推断主导模型，至关重要。此外，主导模型目标参数的置信区间（CI）需要适当考虑站点选择变异性。现有的大多数统计推断方法都依赖于将

合格站点与不合格站点完全分离，而这在实际应用中可能并不现实。目前，能够实现高效且稳健的推断，同时适用于广泛场景，且不依赖完全分离等严格假设的主导模型统计推断方法仍然十分匮乏。

本文通过为利用多源数据对未知主导模型进行联邦元学习推断（RIFL），构建一个具有广泛适用性且有理论依据的框架，填补了这一空白。RIFL方法通过评估各站点在关注参数上的差异，选出用于推断主导模型的合格站点。我们采用一种新颖的重抽样方法来构造具有一致有效性的置信区间。RIFL推断方法对多数群体所属站点与其余站点划分中的错误具有稳健性；参见定理 2。我们还在定理3中证明，在已知合格站点这一先验知识的情况下，我们提出的抽样置信区间可以与预言机置信区间一样短。我们的通用抽样算法能够保护隐私，因为它使用站点特定的汇总统计量实施，且无需在不同站点之间共享个体层面数据。我们提出的RIFL方法通过三个推断问题加以展示：低维参数模型聚合、高维预测模型构建，以及平均处理效应（ATE）的推断。

据我们所知，我们提出的 RIFL 方法是首个能够在多数规则下保证对主导模型具有均匀覆盖率的置信区间。我们还通过模拟研究，将其与另外三种在多数规则下可能使用的推断程序进行了比较，包括多数投票估计量、中位数估计量 [例如 Bowden 等、 2016], 以及 m-out-of-n 自助法 [例如 Chakraborty等人、2013年、Andrews， 2000]。数值结果表明，这三种置信区间均无法达到预期的覆盖性质，而我们的 RIFL 方法则能得到均匀有效的置信区间。我们在第2.3节和第2.4节中解释了这些现有方法覆盖不足的原因。

1.1. 相关文献

RIFL 方法与多个领域的文献相关，包括后选择推断、孟德尔随机化、多源数据整合分析、迁移学习和联邦学习。接下来，我们将详细说明 RIFL 与现有文献的不同之处，并突出其贡献。

为解决后选择推断问题，人们提出了大量新方法和新理论 [例如，Berk等人，2013年，Lee 等人，2016年，Leeb 和 P¨otscher，2005年，Zhang and Zhang，2014年，Javanmard and Montanari，2014年，van de Geer 等，2014年，Chernozhukov 等，2015年，Belloni等人，2014年，Cai and Guo，2017年，Xie和Wang， 2022]。然而，大多数后选择推断文献关注的是高维回归模型中选择少数重要变量后的推断。RIFL框架下的选择问题有着本质区别：选择误差源于不同站点之间的比较，而且没有结果变量来监督选择过程。此外，RIFL只要求满足多数规则，而变量选择方法通常要求只有较小比例的变量会影响结局。

在孟德尔随机化中，人们开发了多种方法，利用多数规则对一维因果效应进行推断 [Bowden 等，2016年，Windmeijer 等人，2019年，Kang 等，2016年，Guo 等人，2018年，Windmeijer 等人， 2021]。Guo[2021] 近期的一项研究证明，工具变量选择错误会导致后选择问题。然而，Guo [2021] 提出的一致有效推断方法依赖于对因果效应一维空间的搜索，

而且难以推广到多变量情形，更不用说高维情形了。相比之下，RIFL 不同于现有的搜索方法，可用于解决范围更广的后选择问题，详情见第5节。

多源数据的整合分析已从不同方向展开研究。Wang等人 [2021] 研究了对有偏来源具有稳健性的数据融合问题。Wang等人 [2021] 的识别条件不同于多数规则，且其方法要有效，必须正确识别无偏来源。Maity等人 [2022]研究了高维情形下的荟萃分析，其中各数据来源相似但并不完全相同；该方法同样要求满足多数规则，并要求多数来源与异常值来源之间存在较大的分离，才能准确识别合格站点。相比之下，RIFL置信区间有效，无需选择步骤准确识别合格站点。Meinshausen和B¨uhlmann [2015], B¨uhlmann和Meinshausen[2015],Rothenh¨ausler等人 [2016], Guo [2020] 针对极大极小效应进行了推断，该效应被定义为适用于异质数据集的稳健预测模型。Cai 等 [2021b], Liu等人[2021], Zhao等人 [2016] 在不同数据来源之间施加了某些相似结构，并对回归模型的共享部分进行了推断。Peters等人 [2016], Arjovsky等人[2019] 研究了多源数据问题，并基于不变性原理识别因果效应。与现有方法不同，RIFL框架仅假设多数站点具有相似模型，同时允许不合格站点与多数群体存在任意差异。

RIFL 与联邦学习和迁移学习领域的现有文献相关。近年来，研究人员开发了保护隐私且通信高效的算法，用于从多个电子健康记录（EHR）来源中学习 [Rasmy等人，2018年，Tong等人， 2022] 以及从多个不同的遗传数据来源中学习 [Kraft等人，2009年，Keys等人， 2020]。联邦回归与预测建模 [Chen等人，2006年，李等人，2013年，Chen和Xie，2014年，Lee等人，2017年，Lian和Fan，2017年，Wang等人，2019年，Duan等人， 2020] 以及因果建模 [Xiong等人，2021年，Vo等人，2021年，Han等人， 2021] 也已得到发展。然而，这些联邦学习方法都没有研究在某些站点可能不具有效性的情况下，如何对主导模型进行推断，而这正是 RIFL 的主要关注点。RIFL 框架也不同于近期开发的迁移学习算法 [，例如 Li等人，2020年，Tian和Feng，2022年，Han等人， 2021]。这些算法要求预先指定一个锚定模型，以便将从源数据集获得的模型与之进行比较。相比之下，RIFL 面向一个更具挑战性的场景：我们不假设存在这样的锚定模型，而是利用多数规则来识别未知的主导模型。

1.2. 论文组织与符号说明

本文的组织结构如下。第 2 节介绍多源数据设置，并强调推断主导模型所面临的挑战。第 3 节提出RIFL方法，第 4 节建立其相关理论。在第5节中，我们通过三个应用示例阐述所提出的方法。在第 6 节中，我们提供了广泛的模拟结果，将我们的方法与现有方法进行比较。第 7 节利用来自四个国家的 16 个参与医疗中心所代表的 275 家医院的真实世界国际电子健康记录数据，介绍我们的方法；这些数据是基于电子健康记录的 COVID-19 临床特征联盟（4CE）多机构联盟的一部分 [Brat 等人， 2020]。

我们介绍本文中使用的记号。对于集合 A， $|A|$ 表示该集合的基数。对于向量 x，我们将其 ℓ_q 范数定义为：对于满足 $\|x\|_0 = \{1 \le l \le p$ 的 $q \ge 0$， $\|x\|_q = \left(\sum_{l=1}^{p} |x_l|^q\right)^{\frac{1}{q}}$ 为 $x_l \ne 0\}|$，且 $\|x\|_\infty =$ 为 $\max_{1\le l\le p} |x_l|$。对于矩阵 X， $X_{i,\cdot}$ 和 $X_{\cdot,j}$ 分别表示其第 i 行和第 j 列。对于两个正数序列 a_n 和 b_n，若 $\limsup_{n\to\infty} a_n/b_n =$0，则 $a_n \ll b_n$。对于矩阵 A，我们用 $\|A\|_F$、 $\|A\|_2$ 和 $\|A\|_\infty$ 分别表示其弗罗贝尼乌斯范数、谱范数，

和逐元素最大范数。

2. 形式化表述与统计推断挑战

2.1. RIFL 的模型假设与概述

本文中，我们假设可以获得从 L 个源总体中抽取的 L 个相互独立的训练数据集。对于 $1 \le l \le L$，我们用 $\mathbb{P}^{(l)}$ 表示第 l 个源总体的分布，并用 $\theta^{(l)} = \theta(\mathbb{P}^{(l)}) \in \mathbb{R}^d$ 表示相应的模型参数。对于任意 $\theta \in \mathbb{R}^d$，我们定义索引集 $\mathcal{V}(\theta) \subset \{1, \cdots, L\}$ 如下：

$$\mathcal{V}(\theta) := \{1 \le l \le L : \theta^{(l)} = \theta\}, \tag{1}$$

其中包含所有具有与 θ. 相同模型参数的站点的索引。现在我们介绍多数规则。

Assumption 1 (Majority Rule). 存在 $\theta^* \in \mathbb{R}^d$，使得 $|\mathcal{V}(\theta^*)| > L/2$.

我们将 θ^* 称为主导模型，它与 $\{\theta^{(l)}\}_{1 \le l \le L}$ 中超过一半的元素相匹配；相应的索引集 $\mathcal{V}(\theta^*)$ 称为主导集合。我们的目标是为 θ^* 的一个低维泛函构造置信域，记为 $\beta^* = g(\theta^*) \in \mathbb{R}^q$，其中 $q \ge 1$, $g(\cdot) \in \mathbb{R}^q$ 是预先指定的低维变换。 $\beta^* = g(\theta^*)$ 的例子包括

(a) Single coefficient or sub-vector: $\beta^* = \theta^*_j$ for $1 \le j \le d$ or $\beta^* = \theta^*_G$ with $G \subset \{1, \cdots, d\}$;

(b) 线性变换：对于任意 $x \in \mathbb{R}^d$，$\intercal \beta^* = x\ \theta^*$；

(c) 二次型： $\beta^* = \|\theta^*\|_2^2$。

为便于记号 起见，我们主要关注 $q = 1$，并在第 3.2 节讨论扩展到具有 $q \ge 2$ 的情形。

若已知主导集合 $\mathcal{V}(\theta^*)$，则可使用标准的元学习和联邦学习方法，利用属于 $\mathcal{V}(\theta^*)$ 的站点数据，对 θ^* 进行推断。然而，正如第 2.3 节所强调的，除多数规则外，在不了解 $\mathcal{V}(\theta^*)$ 的先验知识的情况下，对 θ^* 进行推断要困难得多，因为还需要估计 $\mathcal{V}(\theta^*)$。我们提出的 RIFL 程序包括几个关键步骤：(i) 对于 $l = 1, ..., L$，构造 $\theta^{(l)}$ 和 $\beta^{(l)} = g(\theta^{(l)})$ 的局部估计，分别记为 $\widehat{\theta}^{(l)}$ 和 $\widehat{\beta}^{(l)}$；(ii) 对于 $1 \le l < k \le L$，估计成对差异度量 $\mathcal{D}_{l,k} = D(\theta^{(l)}, \theta^{(k)})$ 以及 $\mathcal{L}_{l,k} = \beta^{(l)} - \beta^{(k)}$，分别记为 $\widehat{\mathcal{D}}_{l,k}$ 和 $\widehat{\mathcal{L}}_{l,k}$，并计算它们的标准误 $\widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,k})$ 和 $\widehat{\text{SE}}(\mathcal{L}_{l,k})$；(iii) 构造主导集合 $\mathcal{V}(\theta^*)$ 的稳健估计；(iv) 推导稳健的基于重抽样的置信集，以 β^*，同时考虑后选择不确定性。 $\widehat{\theta}^{(l)}$ 和 $\widehat{\beta}^{(l)}$ 的构造遵循针对具体研究问题的标准程序。接下来，我们将详细介绍 (ii) 差异度量的构造以及 (iii) 主导集合估计量。RIFL 最具挑战性的步骤是基于重抽样的推断，将在第 3 节中介绍。

2.2. 差异度量

应用多数规则的一个关键步骤，是评估参数 $\theta^{(l)}$ 与 $\theta^{(k)}$ 之间的（差异）度，以用于 $1 \le l,\ k \le L.$。我们构造两组差异度量： $\beta^{(l)}$ 与 $\beta^{(k)}$之间的局部差异度， $\mathcal{L}_{l,\ k} = \beta^{(l)} - \beta^{(k)}$，以及全局差异 $\mathcal{D}_{l,\ k} = \mathcal{D}(\theta^{(l)},\ \theta^{(k)}) =$

$\|\theta^{(l)}-\theta^{(k)}\|_2^2$。尽管也可以为 $\mathcal{D}(\cdot,\ \cdot)$考虑其他向量范数，但我们关注二次范数，因为它具有光滑性且便于进行推断，尤其是在高维情形下。

我们假设 $\{\quad\widehat{\beta}^{(l)},\ \widehat{\sigma}_l\}_{1\le l\le L}$ 满足

$$\frac{1}{\sigma_l}(\widehat{\beta}^{(l)}-\beta^{(l)})\xrightarrow{d}N(0,1)\quad\text{and}\quad\frac{\widehat{\sigma}_l}{\sigma_l}\xrightarrow{p}1, \tag{2}$$

其中 σ_l 表示 $\widehat{\beta}^{(l)}.$ 的标准误。在低维情形下，大多数现有估计量在标准正则条件下满足（2）。在高维情形下，近期已提出各种渐近正态的去偏估计量，并证明其满足（2）；更多讨论请参见第5.2节末尾。

设 $\widehat{\mathcal{L}}_{l,\ k}=\widehat{\beta}^{(l)}-\widehat{\beta}^{(k)}$ 和 $\widehat{\mathcal{D}}_{l,\ k}$ 分别为 $\mathcal{L}_{l,\ k}$ 和 $\mathcal{D}_{l,\ k}$的点估计量。我们将它们的标准误估计为 $\widehat{\text{SE}}(\widehat{\mathcal{L}}_{l,\ k})=\sqrt{\widehat{\sigma}_l^2+\widehat{\sigma}_k^2}$ 和 $\widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,\ k})$，其中 $\widehat{\sigma}_l^2$表示 $\widehat{\beta}^{(l)}$ 的估计方差。对于全局差异度量，我们假设 $\widehat{\mathcal{D}}_{l,\ k}$ 和 $\widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,\ k})$ 满足

$$\limsup_{n\to\infty}\mathbf{P}\left(\left|\widehat{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}\right|/\widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,k})\ge z_\alpha\right)\le\alpha\quad\text{for}\quad 0<\alpha<1, \tag{3}$$

其中，z_α 表示标准正态分布的 α 上分位数。虽然在低维情形下可以将$\mathcal{D}_{l,k}$ 构造为 $\|\widehat{\theta}^{(l)}-\widehat{\theta}^{(k)}\|_2^2$，但由于正则化估计量固有的偏差，在高维情形下推导满足 (3) 的 $\{\widehat{\mathcal{D}}_{l,k},\ \widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,k})\}$ 要困难得多。在第5.1节和第5.2节中，我们证明，对于涵盖低维和高维情形的广泛应用，我们提出的$\mathcal{D}_{l,k}$ 估计量均满足 (3)。

基于这两组差异度量，我们通过以下检验统计量，确定站点 k 与l 在针对 $\beta^*=g(\theta^*)$进行推断时的一致性

$$\widehat{S}_{l,k}:=\max\left\{\left|\widehat{\mathcal{D}}_{l,k}/\widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,k})\right|,\left|\widehat{\mathcal{L}}_{l,k}/\widehat{\text{SE}}(\widehat{\mathcal{L}}_{l,k})\right|\right\}. \tag{4}$$

对于 $1\le l<k\le L$，随后我们可以进行以下显著性检验，以判断第 k个站点和第l个站点是否共享相同的参数，

$$\widehat{H}_{l,k}=\mathbf{1}\left(\widehat{S}_{l,k}\le z_{0.05/[2L(L-1)]}\right), \tag{5}$$

其中，0.05 是预先选定的显著性水平，用于检验不同站点之间的相似性；$z_{0.05/[2L(L-1)]}$ 表示标准正态分布的 $0.05/[2L$（$L-1$）$]$ 上分位数。统计量 $S_{l,k}$ 根据观测数据衡量两个站点彼此不同的证据强度。式（5）中的二元决策 $\widehat{H}_{l,k}$ 本质上估计的是 $H_{l,k}=\mathbf{1}\{\theta^{(l)}=\theta^{(k)}\}.$。我们将阈值设为 $z_{0.05/[2L(L-1)]}$，以调整假设检验的多重性。由于矩阵 H 是对称的，且 $H_{l,l}=1$，因此我们通过令 $\widehat{H}_{k,l}=\widehat{H}_{l,k}$ 以及 $\widehat{H}_{l,l}=1$，构造完整投票矩阵 $\widehat{H}=[\widehat{H}_{l,k}]^{k=1,\ldots,l=1\ \ldots\ L}$，的估计。估计得到的投票矩阵 $\widehat{H}$ 汇总了所有跨站点相似性，并用于估计主导集合。

Remark 1 (Univariate case). 对于单变量 θ^*，这一特殊情形，我们可以简化式 (4) 中检验统计量的构造以及式 (5) 中的投票，得到

$$\widehat{H}_{l,k}=\mathbf{1}\left(\widehat{S}_{l,k}\le z_{0.05/[L(L-1)]}\right)\quad\text{with}\quad\widehat{S}_{l,k}=\left|\widehat{\mathcal{L}}_{l,k}/\widehat{\text{SE}}(\widehat{\mathcal{L}}_{l,k})\right|\quad\text{for}\quad 1\le l<k\le L. \tag{6}$$

2.3. 主导集合估计与后选择问题

接下来，我们构造主导集合 $\mathcal{V}(\theta^*)$ 的两个估计量，并用它们对 β^*. 进行推断。第一个估计量构造如下：

$$\widetilde{\mathcal{V}} := \{1 \le l \le L : \|\widehat{H}_{l,\cdot}\|_0 > L/2\}. \tag{7}$$

集合 $\mathcal{V}$ 包含所有获得“多数票”的站点。第二个估计量是利用图论中的最大团构造的 [Carraghan 和 Pardalos, 1990]。具体而言，我们定义图 $\mathcal{G}([L], H)$，其顶点为 $[L] := \{1, 2, \cdots, L\}$，邻接矩阵为 H，其中 $H_{l,k} = 1$ 和 $H_{l,k} = 0$ 分别表示第 l 个和第 k 个顶点相连和不相连。图 $\mathcal{G}([L], H)$ 的最大团定义为完全连通的最大子图。我们用术语最大团集 表示最大团中对应的顶点集，记作 $\mathcal{MC}([L], H)$。我们通过识别图 $\mathcal{G}([L], H)$ 的最大团集来构造 $\mathcal{V}$，即

$$\widehat{\mathcal{V}} := \mathcal{MC}([L], \widehat{H}). \tag{8}$$

如果多数规则成立，主导集合 $\mathcal{V}(\theta^*)$ 就是 $\mathcal{G}([L], H)$ 的最大团集，也就是说，$\mathcal{V}(\theta^*) = \mathcal{MC}([L], H)$。如果 H 是 H 的足够准确的估计量，那么两个集合估计量 $\mathcal{V}$ 和 $\widehat{\mathcal{V}}$ 都能准确恢复主导集合 $\mathcal{V}(\theta^*)$。然而，由于实践中样本量有限，H 可能不同于真实的 H，因此 $\mathcal{V}$ 和 $\widehat{\mathcal{V}}$ 可能不同于 $\mathcal{V}(\theta^*)$。当最大团集的基数大于 $L/2$, 时，我们有 $\widehat{\mathcal{V}} \subset \mathcal{V}$，也就是说，$\widehat{\mathcal{V}}$ 可能是比 $\mathcal{V}$. 更严格的集合估计量。我们在图1 中说明了式（7）中的 $\mathcal{V}$ 和式（8）中的 $\mathcal{V}$ 的定义。

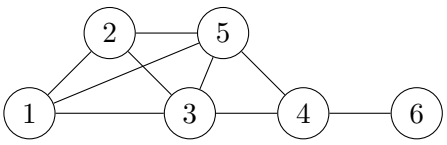

图 1: 图 $\mathcal{G}([L], H)$，包含 $L = 6$. $\mathcal{V} = \{1, 2, 3, 4, 5\}$ 和 $\widehat{\mathcal{V}} = \{1, 2, 3, 5\}$。

接下来，我们展示获得 $\widehat{\mathcal{V}}$. 之后的后续分析。该论证很容易推广到估计集 $\widetilde{\mathcal{V}}$. 。可以聚合 $\{\widehat{\beta}^{(l)}, \widehat{\sigma}_l\}_{l \in \widehat{\mathcal{V}}}$ 来估计 β^*，如下所示的逆方差加权估计量：

$$\widehat{\beta}^* = \frac{\sum_{l \in \widehat{\mathcal{V}}} \widehat{\beta}^{(l)} / \widehat{\sigma}_l^2}{\sum_{l \in \widehat{\mathcal{V}}} 1/\widehat{\sigma}_l^2}. \tag{9}$$

可以为 β^* 构造一个朴素的 $1 - \alpha$ 置信区间：

$$\mathrm{CI}_{\mathrm{post}} = \left(\widehat{\beta}^* - z_{\alpha/2} \frac{1}{\sqrt{\sum_{l \in \widehat{\mathcal{V}}} 1/\widehat{\sigma}_l^2}}, \widehat{\beta}^* + z_{\alpha/2} \frac{1}{\sqrt{\sum_{l \in \widehat{\mathcal{V}}} 1/\widehat{\sigma}_l^2}} \right), \tag{10}$$

其中，$z_{\alpha/2}$ 表示标准正态分布的 $\alpha/2$上分位数。

遗憾的是，与“后选择”文献中的其他情形类似，这种朴素的构造忽略了选择 $\mathcal{V}$时的变异性，可能导致点估计产生偏差，且置信区间覆盖不足。下面通过一个例子说明式（10）中的朴素置信区间所存在的后选择问题。

EXAMPLE 1. 我们针对第 5.3 节所详述的多源因果推断情形，构造目标总体平均处理效应（ATE）的置信区间。我们有 $L = 10$个来源站点，其 $n_l = 1000, 1 \leq l \leq 10$。在每个来源站点，我们观测到数据 $\{X_i^{(l)}, A_i^{(l)}, Y^{(l)}{}_i\}_{1\leq i\leq n_l}$，其中 $X_i^{(l)} \in \mathbb{R}^{10}$ 表示一个 10 维基线协变量向量，$A_i^{(l)} \in \{0, 1\}$ 表示治疗分配（治疗或对照），$Y_i^{(l)} \in \mathbb{R}$ 表示结局。前六个来源站点的生成方式使得目标平均处理效应为 -1，其余四个来源站点的生成方式则使得目标平均处理效应分别为 -1.2、-1.2、-1.1 和 -1.1。在这种情况下，前六个来源站点构成多数群体。依赖于 $\asymp \sim\mathcal{V}$ 和 $\mathcal{V}$ 的置信区间会因纳入了错误选择的站点而出现覆盖不足。根据 500 次模拟，式（10）中的置信区间经验覆盖率仅为 43.2%。如果将式（10）中的$\asymp \sim\mathcal{V}$ 替换为式（7）中的 $\mathcal{V}$，经验覆盖率会降至 27.4%。

2.4. 基于中位数的置信区间面临的挑战

多数规则下，主导模型参数的一种常用相合估计量是中位数估计量 [，例如 Bowden 等、 2016]。我们将中位数估计量构造为 $\{\widehat{\beta}^{(l)}\}_{1\leq l\leq L}$ 的中位数，并通过参数自助法估计其标准误。值得注意的是，尽管当每个站点的样本量趋于无穷时，中位数估计量在多数规则下是相合的，但由于其偏差，它可能不适用于统计推断。因此，基于中位数估计量的置信区间无法达到期望的覆盖性质。为说明这一点，考虑一种特殊情形： L 为奇数，且主导集合中有（$L+1$）/2 个站点。不失一般性，假设 $\mathcal{V}$（θ^*）$= \{1, \ldots,$（$L+1$）$/2\}$。此外，假设 $\beta^{(l)} < \beta^*$ 对于 $l \notin \mathcal{V}$。在此情形下，当非多数站点中的参数值与主导集合中的参数值相距较远时，以高概率，$\widehat{\quad}$ $\widehat{\quad}$中位数估计量将与 min $\{\beta_1, \ldots, \beta_{(L+1)/2}\}$一致。也就是说，中位数估计量是（$\widehat{\beta}_1, \ldots, \widehat{\beta}_{(L+1)/2}$）的最小次序统计量。即使站点特定估计量 $\widehat{\beta}^{(l)}$ 对 β^* 是无偏的，并且在 $l \in \mathcal{V}$下服从正态分布，最小次序统计量通常也不服从正态分布，其均值低于 β^*。更一般地说，中位数估计量的极限分布是一个次序统计量的分布，并且具有对于统计推断而言不可忽略的渐近偏差。Windmeijer 等人在 [2019] 关于无效工具变量的研究中更详细地讨论了这一问题。我们在第 6 节中的数值结果表明，基于中位数估计量的置信区间无法达到期望的覆盖率。

3. RIFL推断

本节中，我们设计基于重抽样的方法，以推导出适用于 β^* 的有效置信区间，从而解决聚合多源数据时的选择后问题。

3.1. RIFL：基于重抽样的推断

RIFL区间构造分为两个步骤。第一步，我们对差异度量进行重抽样，并筛除不准确的重抽样度量。第二步，我们使用重抽样不相似性度量来估计主导集合，并进一步利用该集合生成抽样置信区间。

步骤 1：重抽样与筛选。在给定观测数据的条件下，对于 $1 \le l < k \le L$，我们按照以下方式生成 $\{\widehat{\mathcal{D}}_{l,k}^{[m]}\}_{1\le m\le M}$和 $\{\widehat{\mathcal{L}}_{l,k}^{[m]}\}_{1\le m\le M}$

$$\widehat{\mathcal{D}}_{l,k}^{[m]} \overset{i.i.d}{\sim} N\left(\widehat{\mathcal{D}}_{l,k}, \widehat{\mathrm{SE}}^2(\widehat{\mathcal{D}}_{l,k})\right), \quad \widehat{\mathcal{L}}_{l,k}^{[m]} \overset{i.i.d}{\sim} N\left(\widehat{\mathcal{L}}_{l,k}, \widehat{\mathrm{SE}}^2(\widehat{\mathcal{L}}_{l,k})\right) \quad \text{for} \quad 1 \le m \le M. \tag{11}$$

上述式 (11) 中的生成机制保证了 $\widehat{\mathcal{D}}_{l}^{[m]}{}_{,k} - \widehat{\mathcal{D}}_{l,k}$ 和$\widehat{\mathcal{L}}_{l}^{[m]}{}_{,k} - \widehat{\mathcal{L}}_{l,k}$的分布分别近似于 $\widehat{\mathcal{D}}_{l,k} - \mathcal{D}_{l,k}$和 $\widehat{\mathcal{L}}_{l,k} - \mathcal{L}_{l,k}$的分布。

Remark 2. 随机变量 $\{\widehat{\mathcal{D}}_{l,k}^{[m]}\}_{1\le l<k\le L}$和 $\{\widehat{\mathcal{L}}_{l,k}^{[m]}\}_{1\le l<k\le L}$是独立生成的。即使 $\{\mathcal{D}_{l,k}\}_{1\le l<k\le L}$和 $\{\mathcal{L}_{l,k}\}_{1\le l<k\le L}$ 存在相关性，这种重抽样方法仍然有效。我们提出的方法可以推广，以捕捉 $\{\mathcal{D}_{l,k}\}_{1\le l<k\le L}$和 $\{\mathcal{L}_{l,k}\}_{1\le l<k\le L}$之间的相关结构。不过，我们关注式 (11) 中的重抽样，因为它不要求所有站点都提供所有差异度量的相关结构。

利用重采样数据，我们仿照 (4) 定义重采样检验统计量

$$\widehat{S}_{l,k}^{[m]} := \max\left\{\left|\widehat{\mathcal{D}}_{l,k}^{[m]}/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\right|, \left|\widehat{\mathcal{L}}_{l,k}^{[m]}/\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})\right|\right\} \quad \text{for} \quad 1 \le m \le M. \tag{12}$$

为恰当考虑式 (5) 中定义的检验$H_{l,k}$ 的不确定性，我们证明存在重抽样差异度量 $\{\widehat{\mathcal{D}}_{l,k}^{[m^*]}\}_{1\le l<k\le L}$和 $\{\widehat{\mathcal{L}}_{l,k}^{[m^*]}\}_{1\le l<k\le L}$，它们与相应的真实差异度量 $\{\mathcal{D}_{l,k}\}_{1\le l<k\le L}$ 和 $\{\mathcal{L}_{l,k}\}_{1\le l<k\le L}$ 几乎相同。特别地，以下定理 1 证明，以大于 $1-\nu$的概率，存在 $1 \le m^* \le M$和 $\rho(M) = c_*(\nu)(\log\ n/M)^{1/[L(L-1)]}$，使得

$$\max_{1\le l<k\le L} \max\left\{\left|\frac{\widehat{\mathcal{D}}_{l,k}^{[m^*]} - \mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right|, \left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]} - \mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\right\} \le \rho(M)\cdot T, \quad \text{with} \quad T = z_{\nu/[2L(L-1)]}, \tag{13}$$

其中 $n = \min_{1\le l\le L} n_l$，且 $c_*(\nu)$是依赖于概率 ν、但独立于 n和 d的正数常量。当重抽样规模 M 较大时，我们有 $\rho(M) \to 0$，这表明 $\widehat{\mathcal{D}}^{[m^*]}$和$\widehat{\mathcal{L}}^{[m^*]}$ 分别与 $\mathcal{D}$ 和 $\mathcal{L}$几乎相同。与 (5) 类似，式 (13) 中的阈值水平 T 可解释为邦费罗尼校正阈值水平，用于校正对所有站点之间相似性进行检验时的多重检验。

根据公式 (13)，在构建重采样投票矩阵时，我们将阈值 T 调整为原来的 $\rho(M)$ 倍。具体而言，我们将 $\widehat{H}^{[m]} = \{\widehat{H}_{l,k}^{[m]}\}_{1\le l<k\le L}$ 定义为

$$\widehat{H}_{l,k}^{[m]} = \mathbf{1}\left(\widehat{S}_{l,k}^{[m]} \le \rho(M)\cdot T\right) \quad \text{for} \quad 1 \le l < k \le L. \tag{14}$$

我们定义 $\widehat{H}_{l}^{[m]}{}_{,l} = 1$（当 $1 \le l \le L$，且 $\widehat{H}_{l}^{[m]}{}_{,k} = \widehat{H}_{kl}^{[m]}{}_{,}$时）为 $1 \le k < l \le L.$。关于如何选择调优参数 $\rho(M) \in (0, 1)$ 的经验指导见下文第3.3节。

收缩参数 $\rho(M)$ 降低了式 (14) 中所有成对相似性检验的临界值，因此在判定任意两个站点相似时采用了更严格的规则。式 (13) 中的概率陈述表明，至少有一个重采样投票矩阵能够准确反映真实的

投票矩阵 H，其阈值大幅降低为 $\rho(M) \cdot T$。另一方面，某些重采样统计量 $\widehat{S}_{l\,k}^{[m]}$，可能偏离 0，导致得到一个非常稀疏的 $\widehat{H}^{[m]}$，它可能不同于真实的 H。为了综合利用 M 个重抽样样本的信息，我们首先推断出能够代表底层结构的重抽样样本子集，随后对基于这些合理站点子集构建的置信区间取并集。为此，我们首先计算图 $\mathcal{G}([L],\ \widehat{H}^{[m]})$，的最大团，记为

$$\widehat{\mathcal{V}}^{[m]} = \mathcal{MC}([L], \widehat{H}^{[m]}) \quad \text{for} \quad 1 \le m \le M. \tag{15}$$

我们使用 $|\widehat{\mathcal{V}}^{[m]}|$ 来确定 $\widehat{H}^{[m]}$ 是否与 H 相似。如果 $|\widehat{\mathcal{V}}^{[m]}|$ 小于或等于 $L/2$，我们认为 $\widehat{H}^{[m]}$ 与 H 相差甚远，并舍弃第 m 个重抽样差异度量。我们仅保留属于索引集 $\mathcal{M}$ 的重抽样投票矩阵 $H^{[m]}$，该索引集定义如下

$$\mathcal{M} \coloneqq \left\{1 \le m \le M : |\widehat{\mathcal{V}}^{[m]}| > L/2\right\}. \tag{16}$$

步骤 2：聚合。对于 $m \in \mathcal{M}$，我们构造估计的主导集合 $\widetilde{\mathcal{V}}^{[m]}$，如下所示：

$$\widetilde{\mathcal{V}}^{[m]} = \{1 \le l \le L : \|\widehat{H}_{l,\cdot}^{[m]}\|_0 > L/2\}. \tag{17}$$

集合 $\widetilde{\mathcal{V}}^{[m]}$ 包含所有获得超过半数投票的索引。结合式（17）中的 $\widetilde{\mathcal{V}}^{[m]}$，我们采用逆方差加权估计量

$$\widehat{\beta}^{[m]} = \frac{\sum_{l \in \widetilde{\mathcal{V}}^{[m]}} \widehat{\beta}^{(l)}/\widehat{\sigma}_l^2}{\sum_{l \in \widetilde{\mathcal{V}}^{[m]}} 1/\widehat{\sigma}_l^2}.$$

对于显著性水平 α，我们构造 $1-\alpha$ 置信区间如下：

$$\mathrm{CI}^{[m]} = \left(\widehat{\beta}^{[m]} - z_{\alpha_1/2} \frac{1}{\sqrt{\sum_{l \in \widetilde{\mathcal{V}}^{[m]}} 1/\widehat{\sigma}_l^2}}, \widehat{\beta}^{[m]} + z_{\alpha_1/2} \frac{1}{\sqrt{\sum_{l \in \widetilde{\mathcal{V}}^{[m]}} 1/\widehat{\sigma}_l^2}} \right), \tag{18}$$

其中 $\alpha_1 = \alpha - \nu$，ν 表示预先指定的小概率，用于保证式（13）中的抽样性质。本文中，我们均将 ν 的默认值设为 $\nu = \alpha/20$。

最后，我们将 β^* 的置信区间构造为

$$\mathrm{CI} = \cup_{m \in \mathcal{M}} \mathrm{CI}^{[m]}, \tag{19}$$

其中，$\mathcal{M}$ 和置信区间 $^{[m]}$ 分别在（16）和（18）中定义。尽管 $\cup_{m \in \mathcal{M}}$置信区间$^{[m]}$ 可能并不是一个区间，我们仍将 CI 称为置信区间。用于构造置信区间的 RIFL 算法也总结于补充材料 A.1 节的算法 1 中。

（19）中定义的置信区间的最大值与最小值的平均值可作为 β^*. 的点估计量。此外，我们可以使用 $\widehat{p}_l = \sum_{m \in \mathcal{M}} \mathbf{1}(l \in \widetilde{\mathcal{V}}^{[m]})/|\mathcal{M}|$（即站点 l 被纳入多数群体的次数比例）作为第 l 个站点的可推广性度量。

Remark 3 (Difference between $\widehat{\mathcal{V}}^{[m]}$ and $\widetilde{\mathcal{V}}^{[m]}$). 对于 $m \in \mathcal{M}$，我们有 $\widehat{\mathcal{V}}^{[m]} \subset \widetilde{\mathcal{V}}^{[m]}$，也就是说，式 (17) 中的集合 $\widetilde{\mathcal{V}}^{[m]}$ 与

(16) 中定义的最大团集 $\widetilde{\mathcal{V}}^{[m^*]}$ 相比限制更少。这种关系有助于解释为何我们的构造中使用了两个不同的集合估计量 $\widehat{\mathcal{V}}^{[m]}$ 和 $\widetilde{\mathcal{V}}^{[m]}$。首先，最大团集 $\widehat{\mathcal{V}}^{[m]}$ 的规则比式 (17) 中的 $\widetilde{\mathcal{V}}^{[m]}$ 更严格，而最大团集 $\widehat{\mathcal{V}}^{[m]}$ 可能会筛除更多不准确的重抽样样本。其次，我们的理论表明，存在 $m^* \in \mathcal{M}$，使得 $\widetilde{\mathcal{V}}^{[m^*]}$ 能够保证恢复真实的主导集合 $\mathcal{V}(\theta^*)$。对于 $m \in \mathcal{M}$，集合 $\widetilde{\mathcal{V}}^{[m]}$ 往往包含比 $\widehat{\mathcal{V}}^{[m]}$ 更多的站点，从而使置信区间$^{[m]}$.更短。在式 (17) 中使用 $\widetilde{\mathcal{V}}^{[m]}$ 提高了最终 RIFL 置信区间的精度。

3.2. 扩展至多元目标参数

我们的方法可以轻松扩展，用于为多维目标参数 $\beta^* \in \mathbb{R}^q$ 构造置信区域。作为式 (2) 的推广，我们假设站点特定估计量 $\{\widehat{\beta}^{(l)}, \widehat{\Omega}_l\}_{1 \le l \le L}$ 满足 $\widehat{\Omega}_l^{-1/2}(\widehat{\beta}^{(l)} - \beta^{(l)}) \xrightarrow{d} N(0, I_{q\times q})$，其中 $\widehat{\beta}^{(l)} \in \mathbb{R}^q$，$\widehat{\Omega}_l \in \mathbb{R}^{q\times q}$ 表示 $\widehat{\beta}^{(l)}$ 的估计协方差矩阵，而 $I_{q\times q}$ 是 $q \times q$ 单位矩阵。

我们将重点介绍上述 RIFL 算法针对单变量 β^* 的两项主要调整。首先，我们修改 $\widehat{\mathcal{L}}_{l,\,k}$ 和 $\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,\,k})$ 的计算方式，使其变为 $\widehat{\mathcal{L}}_{l,\,k} = \|\widehat{\beta}^{(l)} - \widehat{\beta}^{(k)}\|_2^2$ 和

$$\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) = \sqrt{4(\widehat{\beta}^{(l)} - \widehat{\beta}^{(k)})^\intercal(\widehat{\Omega}_l + \widehat{\Omega}_k)(\widehat{\beta}^{(l)} - \widehat{\beta}^{(k)}) + 1/\min\{n_l, n_k\}},$$

其中，$1/\min\{n_l, n_k\}$ 用于控制 $\|\widehat{\beta}^{(l)} - \widehat{\beta}^{(k)}\|_2^2$. 的高阶近似误差。其次，我们将方程（18）中置信区间$^{[m]}$ 的构造推广为

$$\mathrm{CS}^{[m]} = \left\{ \beta \in \mathbb{R}^q : (\beta - \widehat{\beta}^{[m]})^\intercal \left(\sum_{l \in \widetilde{\mathcal{V}}^{[m]}} \widehat{\Omega}_l^{-1} \right) (\beta - \widehat{\beta}^{[m]}) \le \chi_q^2(\alpha) \right\} \tag{20}$$

其中，$\widehat{\beta}^{[m]} = \left(\sum_{l\in\widetilde{\mathcal{V}}^{[m]}} \widehat{\Omega}_l^{-1}\right)^{-1} \left(\sum_{l\in\widetilde{\mathcal{V}}^{[m]}} \widehat{\Omega}_l^{-1}\widehat{\beta}^{(l)}\right)$，且 $\chi_q^2(\alpha)$表示具有 q 个自由度的χ^2 分布的上 α 分位数。我们的最终置信集为 $\mathrm{CS} = \cup_{m\in\mathcal{M}}\mathrm{CS}^{[m]}$.

3.3. 调优参数选择

实现 RIFL 置信区间需要指定重抽样规模 M 以及收缩参数 $\rho(M) \in (0, 1)$，后者用于式 (14)。只要规模 M 足够大（例如 $M \ge 500$），所提出的方法对其取值并不敏感。我们将 $M = 500$ 设为默认值。第 4 节的理论结果表明，$\rho(M)$ 的取值可设为 $\rho(M) = c_*(\log n/M)^{1/[L(L-1)]}$，其中正数常量为 c_*；见下文的式 (23)。如果常量 c_* 取值过小，大多数重抽样差异度量都无法生成满足多数规则的最大团集，导致 $|\mathcal{M}|$非常小。因此，我们可以用 $|\mathcal{M}|$ 来判断调优参数 $\rho(M)$ 的取值是否足够大。在实践中，我们从较小的 c_* 值（例如 $c_* = 1/12$）开始，逐步增大 c_*，直到预先指定的比例（例如 $\mathrm{prop} = 10\%$）的重抽样差异度量生成满足多数规则的最大团集。RIFL 置信区间几乎不受式 (13) 中原始阈值 T 的影响。我们所选的 $\rho(M)$ 将确保 $\rho(M) \cdot T$ 足够大，使超过 $\mathrm{prop} = 10\%$ 的重抽样集合满足多数规则。即使阈值 T 因邦费罗尼校正而较为保守，T 的这一取值也不会影响 RIFL 置信区间的表现，因为 $\rho(M)$ 的取值会根据阈值 T. 的设定进行自适应调整。我们在第 6.4 节的数值研究中展示了该方法对调优参数的稳健性。

4. RIFL推断的理论依据

接下来，我们为 RIFL 置信区间的有效性提供理论依据。令 $n = \min_{1\le l\le L} n_l$，并定义

$$\mathrm{err}_n(M,\nu) = c_*(\nu)\left[\frac{\log n}{M}\right]^{\frac{1}{L(L-1)}} \text{ with } c_*(\nu) = 2^{\frac{1}{L(L-1)}-\frac{1}{2}}\sqrt{\pi}\exp\left(\frac{1}{2}z^2_{\nu/[2L(L-1)]}\right), \tag{21}$$

其中 $L \ge 2$ 和 $0 < \nu < 1/2.$。项 $\mathrm{err}_n(M,\nu)$ 量化了抽样精度，表示经过 M 次重抽样后，真实差异度量与重抽样所得度量之间的最小差异。对于常数 $\nu \in (0, 1/2)$ 和固定的 L, $c_*(\nu)$ 是一个常数，仅取决于预先指定的概率 ν，而 $\mathrm{err}_n(M,\nu)$ 的阶为 $(\log n/M)^{\frac{1}{L(L-1)}}$，当 M 足够大时趋于 0。

以下定理确立了关键的抽样性质，为式 (14) 中的阈值降低提供了理论支撑。

THEOREM 1. 假设站点特定估计量 $\{\widehat{\beta}^{(l)}, \widehat{\sigma}_l\}_{1\le l\le L}$ 满足 (2)，且差异度量 $\{\widehat{\mathcal{D}}_{l,k}, \widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\}_{1\le l<k\le L}$ 满足 (3)。那么，式 (11) 中定义的重抽样差异度量 $\{\widehat{\mathcal{D}}^{[m]}_{l,k}\}_{1\le m\le M}$ 和 $\{\widehat{\mathcal{L}}^{[m]}_{l,k}\}_{1\le m\le M}$ 满足

$$\liminf_{n\to\infty}\liminf_{M\to\infty}\mathbf{P}\left(\min_{1\le m\le M}\left[\max_{1\le l<k\le L}\max\left\{\left|\frac{\widehat{\mathcal{D}}^{[m]}_{l,k}-\mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right|, \left|\frac{\widehat{\mathcal{L}}^{[m]}_{l,k}-\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\right\}\right] \le \mathrm{err}_n(M,\nu)\right) \ge 1-\nu,$$

对任意正数常量 $0 < \nu < 1/2.$

上述 定理表明，以较高概率存在 $1 \le m^* \le M$，使得

$$\max_{1\le l<k\le L}\max\left\{\left|\frac{\widehat{\mathcal{D}}^{[m^*]}_{l,k}-\mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right|, \left|\frac{\widehat{\mathcal{L}}^{[m^*]}_{l,k}-\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\right\} \le \mathrm{err}_n(M,\nu) \asymp \left(\frac{\log n}{M}\right)^{\frac{1}{L(L-1)}}.$$

这为式 (13) 提供了理论依据。

以下定理确立了所提出的 RIFL置信区间的覆盖性质。

THEOREM 2. 假设定理 *1* 的假设成立，且式 (13) 中使用的阈值水平 $\rho(M)\cdot T$ 满足

$$\rho(M)\cdot T \ge \mathrm{err}_n(M,\nu) \quad and \quad \lim_{M\to\infty}\rho(M)\cdot T = 0. \tag{22}$$

那么，(19)中定义的置信区间满足

$$\liminf_{n\to\infty}\liminf_{M\to\infty}\mathbf{P}\left(\beta^* \in \mathrm{CI}\right) \ge 1-\alpha,$$

其中，$\beta^* = g(\theta^*)$ 和 $0 < \alpha < 1$ 分别表示(18)中使用的显著性水平。

Eq公式 (22) 是关于收缩参数 $\rho(M)$。我们可以选择 $\rho(M)$ a s

$$\rho(M) = \mathrm{err}_n(M,\nu)/T = c_*(\log n/M)^{\frac{1}{L(L-1)}}, \quad \text{with} \quad c_* = c_*(\nu)/T, \tag{23}$$

其中 c_* 是一个与 n 和 d 无关的常数。当 M 足够大时，(23) 中对 $\rho(M)$ 的选择会自动满足条件 (22)。

下面我们将展示，RIFL方法能够检测出所有其模型参数与主导模型充分分离的站点。因此，当非多数群体中的所有站点都与多数群体中的站点充分分离时，在先验知识 $\mathcal{V}(\theta^*)$ 的前提下，RIFL置信区间与预言机置信区间一致。利用先验知识 $\mathcal{V}^* = \mathcal{V}(\theta^*)$，我们可以构造预言机 $1-\alpha$ 置信区间如下

$$\mathrm{CI}_{\mathrm{ora}} = \left(\widehat{\beta}^{\mathrm{ora}} - z_{\alpha/2} \frac{1}{\sqrt{\sum_{l\in\mathcal{V}^*} 1/\widehat{\sigma}_l^2}}, \widehat{\beta}^{\mathrm{ora}} + z_{\alpha/2} \frac{1}{\sqrt{\sum_{l\in\mathcal{V}^*} 1/\widehat{\sigma}_l^2}} \right), \text{ with } \widehat{\beta}^{\mathrm{ora}} = \frac{\sum_{l\in\mathcal{V}^*} \widehat{\beta}^{(l)}/\widehat{\sigma}_l^2}{\sum_{l\in\mathcal{V}^*} 1/\widehat{\sigma}_l^2}. \tag{24}$$

THEOREM 3. 假设定理 *2* 的条件成立。对于 $k \in \mathcal{V}^c$，若

$$|\beta^{(k)} - \beta^*| \geq (2\sqrt{2\log n + 2\log M} + \rho(M)\cdot T) \cdot \max_{l\in\mathcal{V}} \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}), \tag{25}$$

则 $\lim_{n\to\infty}\lim_{M\to\infty} \mathbf{P}\left(k \notin \cup_{m\in\mathcal{M}} \widetilde{\mathcal{V}}^{[m]}\right) = 1$. 此外，如果 $|\mathcal{V}| = \lfloor L/2 \rfloor + 1$ 且任意满足(25)的 $k \in \mathcal{V}^c$，则(19) 中定义的置信区间满足

$$\lim_{n\to\infty}\lim_{M\to\infty} \mathbf{P}\left(\mathrm{CI} = \mathrm{CI}_{\mathrm{ora}}\right) = 1.$$

条件 (25) 要求非多数站点 k 与多数站点存在显著差异。如此明显分离的非多数站点 k 不会被纳入 (17) 中定义的任何重采样主导集合 $\widetilde{\mathcal{V}}^{[m]}$。当所有非多数站点都容易检测，且超过半数的站点属于 $\mathcal{V}$ 时，上述定理表明，所提出的置信区间可以与已知 $\mathcal{V}^*$ 时的预言机置信区间相同。

5. 多源推断问题的应用

本节详细介绍 RIFL 方法在实践中常见的三类推断问题中的应用。我们将探讨：(i) 第 5.1 节中一般参数化建模框架下的稳健推断；(ii) 第 5.2 节中关于主导高维预测模型的统计推断问题；(iii) 第 5.3 节中基于多个站点估计平均处理效应。在这些广泛的应用中，主导模型备受关注，因为它是一个具有普适性的模型，能够匹配大多数观测数据集。应用 RIFL 方法需要构建站点特定估计量 $\{\widehat{\beta}^{(l)}, \widehat{\sigma}_l\}_{1\leq l\leq L}$，以及差异度量$\frown \quad \frown \frown\{\mathcal{D}_{l,k}$,SE($\mathcal{D}_{l,\ k}$)$\}_{1\leq l<k\leq L}$，以下将针对每种应用详细说明。

5.1. 参数模型推断

我们假设可以访问 L 个站点，其中第 l 个站点（具有 $1 \leq l \leq L$）是一个参数模型，其对应的模型参数为 $\theta^{(l)} \in \mathbb{R}^d$. 。我们在

当前小节中关注低维情形，并将在第5.2节讨论高维模型。对于 $1 \le l \le L$，第 l 个站点输出一个估计量 $\widehat{\theta}^{(l)}$，满足

$$\sqrt{n_l}(\widehat{\theta}^{(l)} - \theta^{(l)}) \xrightarrow{d} N(0, C^{(l)}), \tag{26}$$

其中，协方差矩阵 $C^{(l)} \in \mathbb{R}^{d\times d}$ 由 $\widehat{C}^{(l)}$相合地估计。此一般性设定涵盖许多参数和半参数模型，包括但不限于广义线性模型、生存模型和分位数回归。许多现有估计量，包括 M 估计量，在正则条件下满足条件 (26)。

我们的目标是对预先指定的变换 $g(\theta^*)$，进行推断，其中主导模型 θ^* 定义为与超过一半的 $\{\theta^{(l)}\}_{1\le l\le L}$. 一致的模型。我们用代入估计量 $\widehat{\beta}^{(l)} = g(\widehat{\theta}^{(l)})$估计 $\beta^{(l)}$。对于二阶可微函数 g，我们应用 Delta 方法，并将 $\widehat{\beta}^{(l)}$ 的标准误估计量构造为 $\widehat{\sigma}_l = \sqrt{\left[\nabla g(\widehat{\theta}^{(l)})\right]^{\intercal} \widehat{C}^{(l)} \nabla g(\widehat{\theta}^{(l)})}$，其中 $\nabla g(\cdot)$ 表示函数 g的梯度。我们用二次范数度量两个站点之间的差异度

$$\mathcal{D}_{l,k} = \|\theta^{(l)} - \theta^{(k)}\|_2^2 \quad \text{for} \quad 1 \le l < k \le L.$$

接下来，我们给出估计量 $\mathcal{D}_{l,k}$ 及其标准误估计量 SE($\mathcal{D}_{l,k}$)，然后应用 RIFL 为 $g(\theta^*)$ 构造置信区间。

对于给定的 l和 k，定义 $\gamma_{l,k} = \theta^{(l)} - \theta^{(k)}$ 和 $\widehat{\gamma}_{l,k} = \widehat{\theta}^{(l)} - \widehat{\theta}^{(k)}$. 为简化表述，稍微滥用一下记号，我们在下文的 γ 和 $\widehat{\gamma}$ 中省略下标 $_{l,k}$。我们将 $\widehat{\mathcal{D}}_{l,k} = \|\widehat{\gamma}\|_2^2$ 的误差分解如下:

$$\widehat{\mathcal{D}}_{l,k} - \mathcal{D}_{l,k} = \|\widehat{\gamma}\|_2^2 - \|\gamma\|_2^2 = 2\langle\widehat{\gamma} - \gamma, \gamma\rangle + \|\widehat{\gamma} - \gamma\|_2^2. \tag{27}$$

右侧的第一项 $2\langle\widehat{\gamma} - \gamma, \gamma\rangle$ 满足

$$\frac{2\langle\widehat{\gamma} - \gamma, \gamma\rangle}{\sqrt{4\gamma^{\intercal}C^{(l)}\gamma/n_l + 4\gamma^{\intercal}C^{(k)}\gamma/n_k}} \xrightarrow{d} N(0,1).$$

结合式 (27) 中的分解，我们将 $\mathcal{D}_{l,k}$ 的标准误估计为

$$\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k}) = \sqrt{4\widehat{\gamma}^{\intercal}\widehat{C}^{(l)}\widehat{\gamma}/n_l + 4\widehat{\gamma}^{\intercal}\widehat{C}^{(k)}\widehat{\gamma}/n_k + 1/\min\{n_l, n_k\}}. \tag{28}$$

在 SE($\mathcal{D}_{l,\ k}$) 的定义中加入额外项 $1/\min\{n_l, n_k\}$，是为了控制式（27）右侧第二项 $\|\widehat{\gamma} - \gamma\|_2^2$ 的不确定性。$\widehat{\beta}^{(l)}$，$\widehat{\sigma}_l$，$\widehat{\mathcal{D}}_{l,\ k}$ 以及 SE($\widehat{\mathcal{D}}_{l,\ k}$) 的构造仅依赖汇总统计量 $\{\widehat{\theta}^{(l)}\}_{1\le l\le L}$，因此无需共享个体层面数据即可实现 $\{\widehat{C}^{(l)}\}_{1\le l\le L}$. RIFL。下面的定理 4 说明，式（28）中的 $\widehat{\mathcal{D}}_{l,\ k} = \|\widehat{\gamma}\|_2^2$ 和 SE($\mathcal{D}_{l,\ k}$) 满足 $1 \le l < k \le L$ 的条件（3）。

THEOREM 4. 假设 $\{\widehat{\theta}^{(l)}\}_{1\le l\le L}$ 满足(26)，且 $C^{(l)}$的最大特征值（记为 $\lambda_{\max}(C^{(l)})$）有界。对于 $1 \le l \le L.$，若 $\widehat{C}^{(l)}$ 是 $C^{(l)}$的相合估计量，则,$\widehat{\mathcal{D}}_{lk} = \|\widehat{\gamma}\|_2^2$ 及其标准误估计量SE($\widehat{\mathcal{D}}_{l,k}$) （见(28)）对于 $1 \le l < k \le L$满足(3)。

5.2. 高维预测模型

接下来，我们通过汇总来自 L 个站点的信息，考虑一个常见高维预测模型的推断问题。对于第 l 个站点，我们考虑数据 $\{X_i^{(l)}, Y_{(l)\,i}\}_{1\le i\le n_l}$,对应的以下广义线性模型（GLM）：

$$\mathbf{E}(Y_i^{(l)} \mid X_i^{(l)}) = h(\mu_l + [X_i^{(l)}]^\intercal \theta^{(l)}) \quad \text{for} \quad 1 \le i \le n_l,$$

其中， $h(\cdot)$ 是已知的链接函数， μ_l 是截距， $\theta^{(l)} \in \mathbb{R}^d$ 是与协变量对应的回归向量。我们假设 $\theta^{(l)}$ 在主导集合内共享，但允许 μ_l 在不同站点之间有所差异。为便于说明，对于连续型结局，我们采用恒等链接函数 $h(x) = x$ ；对于二元结局，则采用逻辑链接函数 $h(x) = 1/[1+\exp(-x)]$ 。不过，我们的方案可以扩展到其他链接函数 h。

我们采用与第5.1节低维情形相同的二次差异度量。与低维情形相比，在高维情形下构造 $\mathcal{D}_{l,k} = \|\theta^{(l)} - \theta^{(k)}\|_2^2$ 的准确估计量要困难得多。要确立高维情形下差异度估计量的理论性质，需要进行样本拆分。对于 $1 \le l \le L$，我们将索引集 $\{1, 2, \cdots, n_l\}$ 随机拆分为互不相交子集 $\mathcal{S}_1^{(l)}$ 和 $\mathcal{S}_2^{(l)}$，其中 $|\mathcal{S}_1^{(l)}| = \lceil n_l/2 \rceil$ 和 $|\mathcal{S}_2^{(l)}| = n_l - |\mathcal{S}_1^{(l)}|$，这里 $\lceil n_l/2 \rceil$ 表示大于 $n_l/2.$ 的最小整数。对于任意集合 $\mathcal{S} \subseteq \{1, ..., n_l\}$，令 $\widehat{\mathbb{P}}_{\mathcal{S}}^{(l)} f(Y, X) = |\mathcal{S}|^{-1} \sum_{i\in\mathcal{S}} f(Y_{(l)i}, X_i^{(l)})$。

我们的方案分为三个步骤：第一步，我们构造初始估计量 $\{\widetilde{\mu}_l, \widetilde{\theta}^{(l)}\}_{1\le l\le L}$以及去偏估计量 $\{\widehat{\beta}^{(l)}\}_{1\le l\le L}$ 和相应的标准误 $\{\widehat{\sigma}_l\}_{1\le l\le L}$，并将这些初始估计量广播给所有 L 个站点；第二步，我们估计代入估计量的误差分量；第三步，我们构造$\mathcal{D}_{l,k}$的去偏估计量，用于 $1 \le l < k \le L.$ 。我们的方法旨在保护隐私，因为它不需要传输个体层面数据。

步骤 1：广播初始估计量。对于具有 $1 \le l \le L,$ 的站点 l ，我们使用属于 $\mathcal{S}_1^{(l)}$ 的数据构造惩罚最大似然估计量 [B¨uhlmann 和 van de Geer, 2011] 。具体而言， $\theta^{(l)}$ 的初始估计量定义为

$$\left\{\widetilde{\mu}_l, \widetilde{\theta}^{(l)}\right\} = \underset{\mu\in\mathbb{R}, \theta\in\mathbb{R}^d}{\arg\min} \left[\widehat{\mathbb{P}}_{\mathcal{S}_1^{(l)}}^{(l)} \{\ell_h(\mu + \theta^\intercal X, Y)\} + \lambda\|\theta\|_1\right], \tag{29}$$

其中， $\ell_h(x, y)$对应于对数似然函数。对于具有 $h(x) = x$ 的线性模型，其形式为（$y - x$）^{2}；对于具有 $h(x) = e^x/(1 + e^x)$的逻辑回归模型，其形式为$\log(1 + e^x) - yx$ 。实际应用中，调优参数 $0 < \lambda \asymp \sqrt{\log d/|\mathcal{S}_1^{(l)}|}$ 通过交叉验证选取。

由于我们的最终目标是对主导模型 θ^* 中的 $\beta^* = g(\theta^*)$ 进行推断，因此我们要求所有站点额外输出一个关于 $\beta^{(l)} = g(\theta^{(l)})$ 的去偏渐近正态估计量 $\widehat{\beta}^{(l)}$ 。在高维情形下，代入估计量 $g(\widetilde{\theta}^{(l)})$ 通常是 $\beta^{(l)}$ 的有偏估计量，因此需要进行偏差校正。近年来，这类偏差校正程序得到了广泛研究 [Zhang and Zhang, 2014, Javanmard and Montanari, 2014, van de Geer 等, 2014, Chernozhukov 等, 2015, Ning and Liu, 2017, Cai 等, 2021a], , 其例子包括线性泛函 [Zhu and Bradic, 2018, Cai 等, 2021c, Guo 等人, 2021a], 和二次型 [Guo 等人, 2021b, Cai and Guo, 2020]。我们将通过考虑高维回归模型中某个特定坐标 j 的例子 $\beta^{(l)} = \langle e_j, \theta^{(l)}\rangle = \theta_j^{(l)}$ 来说明我们的方法。遵循 Zhang and Zhang [2014],

Javanmard and Montanari [2014], van de Geer 等 [2014], 针对 $\beta^{(l)}$ 的去偏估计量已被证明满足以下渐近正态性性质:

$$(\widehat{\beta}^{(l)} - \beta^{(l)})/\mathrm{SE}(\widehat{\beta}^{(l)}) \xrightarrow{d} N(0,1) \tag{30}$$

并且还给出了 SE($\widehat{\beta}^{(l)}$) 的相合估计量，记为 $\widehat{\ }$SE($\widehat{\beta}^{(l)}$)。

随后，我们将 $\{\widetilde{\mu}_l, \widetilde{\theta}^{(l)}, \widehat{\beta}^{(l)}, \widehat{\ }\mathrm{SE}(\widehat{\beta}^{(l)})\}_{1\le l\le L}$ 广播给所有 L 站点。

步骤 2：估计代入估计量的偏差分量。在步骤 2 中，我们获取了构造 $\{\mathcal{D}_{l,k}\}_{1\le l<k\le L}$的去偏估计量所需的额外汇总数据。为计算距离度量并评估其在高维情形下的抽样误差，我们需要校正用于 $1 \le l < k \le L$的代入估计量 $\|\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}\|_2^2$ 的偏差。距离度量估计的偏差校正需要为每个成对距离构造一个投影。为此，对于 $1 \le l \le L$，我们定义 $\eta^{(l)} = (\mu_l, [\theta^{(l)}]^\intercal) \in \mathbb{R}^{d+1}$和 $\widetilde{\eta}^{(l)} = (\widetilde{\mu}_l, [\widetilde{\theta}^{(l)}]^\intercal) \in \mathbb{R}^{d+1}$ 以及$\widetilde{X}_i^{(l)} = (1, (X_i^{(l)})^\intercal)^\intercal$，对于 $1 \le i \le n_l.$，我们固定索引 $1 \le l < k \le L$，并令 $\widehat{\gamma}_{l,k} := (0, [\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}]^\intercal)^\intercal$ 和 $\gamma_{l,k} := (0, [\theta^{(l)} - \theta^{(k)}]^\intercal)^\intercal$。根据上下文需要，我们分别将 $\widetilde{\gamma}_{l,k}$和$\gamma_{l,k}$简写为 $\widetilde{\gamma}$和 γ。

我们分析代入估计量$\|\widetilde{\gamma}\|_2^2 = \|\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}\|_2^2.$ 的误差分解。与低维情形的主要区别在于，由于式（29）中的惩罚项，$\widetilde{\theta}^{(l)}$ 是 $\theta^{(l)}$ 的有偏估计量。因此，代入估计量 $\|\widetilde{\gamma}\|_2^2$ 对于 $\|\gamma\|_2^2$ 可能存在严重偏差。为减轻这种过度偏差，我们根据以下分解提出一种偏差校正程序

$$\|\widetilde{\gamma}\|_2^2 - \|\gamma\|_2^2 = 2\langle\widetilde{\eta}^{(l)} - \eta^{(l)}, \widetilde{\gamma}\rangle - 2\langle\widetilde{\eta}^{(k)} - \eta^{(k)}, \widetilde{\gamma}\rangle - \|\widetilde{\gamma} - \gamma\|_2^2. \tag{31}$$

与低维情形不同，项 $2\langle\widetilde{\eta}^{(l)} - \eta^{(l)}, \widetilde{\gamma}\rangle$ 和 $2\langle\widetilde{\eta}^{(k)} - \eta^{(k)}, \widetilde{\gamma}\rangle$ 由于 $\widetilde{\eta}^{(l)}$ 和 $\widetilde{\eta}^{(k)}$ 而存在过度偏差，无法像低维情形那样达到渐近正态极限分布。

遵循 Guo 等人 [2021a], 的方法，我们通过估计误差分量 $2\langle\widetilde{\eta}^{(l)} - \eta^{(l)}, \widetilde{\gamma}\rangle$，并使用属于 $\mathcal{S}_2^{(l)}$ 和 $\mathcal{S}_2^{(k)}.$ 的数据进行去偏。对于 $1 \le i \le n_l$，定义

$$\epsilon_i^{(l)} = Y_i^{(l)} - h([\widetilde{X}_i^{(l)}]^\intercal\eta^{(l)}),\ \widehat{\epsilon}_i^{(l)} = Y_i^{(l)} - h([\widetilde{X}_i^{(l)}]^\intercal\widetilde{\eta}^{(l)}),\ W_i^{(l)} = 1/h'([\widetilde{X}_i^{(l)}]^\intercal\widetilde{\eta}^{(l)}),$$

其中 $h'(x) = dh(x)/dx$。注意,

$$u^\intercal\widehat{\mathbb{P}}^{(l)}_{\mathcal{S}_2^{(l)}}(W\widetilde{X}\widehat{\epsilon}) \approx u^\intercal\widehat{\mathbb{P}}^{(l)}_{\mathcal{S}_2^{(l)}}(W\widetilde{X}\epsilon) + u^\intercal\widehat{\Sigma}^{(l)}(\eta^{(l)} - \widetilde{\eta}^{(l)}) \tag{32}$$

其中 $\widehat{\Sigma}^{(l)} = \widehat{\mathbb{P}}^{(l)}_{\mathcal{S}_2^{(l)}}\{\widetilde{X}(\widetilde{X})^\intercal\}$。根据 (32)，我们旨在构造投影方向 $u \in \mathbb{R}^{d+1}$，使得 $\widehat{\Sigma}^{(l)}u \approx \widetilde{\gamma}_{l,\ k}$，从而确保 $u\ \widehat{\mathbb{P}}^{(l)}_{\mathcal{S}_2^{(l)}}(W\widetilde{X}\widehat{\epsilon})$ 是 $\langle\widetilde{\eta}^{(l)} - \eta^{(l)}, \widetilde{\gamma}\rangle$ 的准确估计量。具体而言，我们记 $\widehat{\gamma}_{l,\ k} = (0, [\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}]^\intercal)^\intercal$，并按如下方式构造 $\widehat{u}_k^{(l)}$，

$$\begin{aligned}\widehat{u}_k^{(l)} = \underset{u\in\mathbb{R}^d}{\arg\min}\, u^\intercal\widehat{\Sigma}^{(l)}u \quad \text{subject to } &\|\widehat{\Sigma}^{(l)}u - \widetilde{\gamma}_{l,k}\|_\infty \le \|\widetilde{\gamma}_{l,k}\|_2\lambda \\ &|\widetilde{\gamma}_{l,k}^\intercal\widehat{\Sigma}^{(l)}u - \|\widetilde{\gamma}_{l,k}\|_2^2| \le \|\widetilde{\gamma}_{l,k}\|_2^2\lambda \\ &\max_{i\in\mathcal{S}_2^{(l)}}|u^\intercal\widetilde{X}_i^{(l)}| \le \|\widetilde{\gamma}_{l,k}\|_2^2\tau\end{aligned} \tag{33}$$

其中 $\lambda \asymp \sqrt{\log d / |\mathcal{S}_2^{(l)}|}$ 和 $\tau \asymp \sqrt{\log |\mathcal{S}_2^{(l)}|}$。接下来，我们得到去偏项

$$\widehat{\delta}_k^{(l)} = \left(\widehat{u}_k^{(l)}\right)^{\mathsf{T}} \widehat{\mathbb{P}}_{\mathcal{S}_2^{(l)}}^{(l)}(W\widetilde{X}\widehat{\epsilon}), \quad \text{for} \quad 1 \le k \le L,\ k \ne l, \tag{34}$$

以及相应的方差度量

$$\widehat{\mathrm{V}}_k^{(l)} = \left(\widehat{u}_k^{(l)}\right)^{\mathsf{T}} \left[\widehat{\mathbb{P}}_{\mathcal{S}_2^{(l)}}^{(l)}(W\widetilde{X}\widetilde{X}^{\mathsf{T}})\right] \widehat{u}_k^{(l)} \quad \text{for} \quad 1 \le k \le L,\ k \ne l. \tag{35}$$

类似地，我们可以获得 $\widehat{u}_l^{(k)}$、$\widehat{\delta}_l^{(k)}$ 和 $\widehat{\mathbf{V}}_l^{(k)}$，用于估计 $\langle \widetilde{\eta}^{(k)} - \eta^{(k)}, \widetilde{\gamma}_{l,k} \rangle$。

步骤 3：构造去偏估计量。在最后一步，中心将所有站点的汇总统计量结合起来，构造以下 $\mathcal{D}_{l,\,k}$ 的偏差校正估计量，

$$\widehat{\mathcal{D}}_{l,k} = \max\left\{ \|\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}\|_2^2 + 2\widehat{\delta}_k^{(l)} - 2\widehat{\delta}_l^{(k)}, 0 \right\}, \tag{36}$$

并通过以下方式估计其标准误

$$\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k}) = \sqrt{4\widehat{\mathrm{V}}_k^{(l)} + 4\widehat{\mathrm{V}}_l^{(k)} + \frac{1}{\min\{n_l, n_k\}}}. \tag{37}$$

我们已在补充材料的算法 2 中总结了所提出的方法。在补充材料的定理 5 中，我们证明了式（36）中的估计量和式（37）中的 $\mathrm{SE}(\mathcal{D}_{l,\,k})$ 满足 $1 \le l < k \le L$ 的条件（3）。利用式（36）和（37）中定义的 $\{\widehat{\beta}^{(l)}, \widehat{\mathrm{SE}}(\widehat{\beta}^{(l)}) \widehat{\mathcal{D}}_{l,\,k}, \widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,\,k})\}_{1 \le l < k \le L}$，我们应用 RIFL 为 β^* 构造置信区间。

5.3. 平均处理效应的推断

多中心因果建模对于生成获批药物在新目标总体中的可靠真实世界证据具有重要价值；该目标总体可能远超临床试验招募的患者范围。由于样本量有限，单个站点的数据可能不足以可靠地估计因果效应；混杂调整不足也可能导致潜在偏差。我们旨在使用 RIFL，基于 L 个来源站点，估计具有指定协变量分布 $f_G(\cdot)$ 的目标总体的平均处理效应。具体而言，对于站点 $1 \le l \le L$，我们观测到数据 $\{X_i^{(l)}, A_i^{(l)}, Y_{(l)\,i}\}_{1 \le i \le n_l}$，其中 $X_i^{(l)} \in \mathbb{R}^p$ 表示 p维基线协变量向量，$A_i^{(l)} \in \{0,1\}$ 表示受试者是否接受试验性治疗或对照，$Y_i^{(l)} \in \mathbb{R}$ 表示结局。此外，对于某个较大的 N，我们观测到从 $f_G(\cdot)$ 中抽取的 $\{X_i^{\mathcal{T}}\}_{1 \le i \le N}$，该分布是目标总体中的协变量分布。令 $f_l(\cdot)$ 表示第 l 个来源站点中协变量分布的密度函数。我们用 $Y_{(l,a)}$ 表示站点 l 中患者接受治疗 $A = a$ 时的潜在结局。我们将由第 l 个来源站点投射得到的目标总体的平均处理效应定义为

$$\theta^{(l)} = \int \left\{ \mathbf{E}(Y^{(l,1)} \mid X^{(l)} = x) - \mathbf{E}(Y^{(l,0)} \mid X^{(l)} = x) \right\} f_G(x) dx.$$

我们使用 β^* 表示大多数平均处理效应 $\{\theta^{(l)}\}_{1 \le l \le L}$，并将 $\mathcal{V}(\beta^*) = \{1 \le l \le L$ 定义为：$\theta^{(l)} = \beta^*\}$。我们旨在使用 RIFL 对 β^* 进行推断。

为估计 β^*，我们首先使用第 l 个站点的数据获得一个标准的增广双重稳健估计量 [（Bang 和 Robins，2005年；Tsiatis，2006年；Kang 和 Schafer，2007年；Tao 和 Fu， 2019]），用于 $\theta^{(l)}$，该估计量需要倾向评分和结局模型

$$\mathbf{P}(A_i^{(l)} = a \mid X_i^{(l)}) = \pi_l(a, X_i^{(l)}; \boldsymbol{\alpha}^{(l)}), \quad \mathbf{E}(Y_i^{(l)} \mid A_i^{(l)} = a, X_i^{(l)}) = m(a, X_i^{(l)}; \boldsymbol{\gamma}_a^{(l)}),$$

其中， $\boldsymbol{\alpha}^{(l)}$ 和 $\{\boldsymbol{\gamma}_0^{(l)}, \boldsymbol{\gamma}_1^{(l)}\}$ 是未知的模型参数， $\pi_l(\cdot)$和 $m(\cdot)$是指定的链接函数。为了推断目标站点平均处理效应，还需要指定一个密度比模型，$f_G(X_i^{(l)})/f_l(X_i^{(l)}) = \omega_l(X_i^{(l)}; \boldsymbol{\eta}^{(l)})$，用于处理源总体与目标总体之间的协变量偏移。 $\omega_l(X_i^{(l)}; \boldsymbol{\eta}^{(l)})$可以例如取为 $\exp\{\Psi(X_i^{(l)})^\intercal \boldsymbol{\eta}^{(l)}\}$，其中$\Psi(\cdot)$表示指定基函数构成的向量。

令 $\widehat{\boldsymbol{\alpha}}^{(l)}$、 $\widehat{\boldsymbol{\gamma}}_a^{(l)}$和 $\widehat{\boldsymbol{\eta}}^{(l)}$ 分别表示 $\boldsymbol{\alpha}^{(l)}$、 $\boldsymbol{\gamma}_a^{(l)}$和 $\boldsymbol{\eta}^{(l)}$的相合估计量。例如，可以在各个站点 l 通过拟合广义线性模型对这些有限维参数进行局部估计。对于满足 $1 \le l \le L$的站点 l，我们计算以下双重稳健估计量 $\widehat{\theta}^{(l)} = \widehat{M}^{(l)} + \widehat{\delta}^{(l)}$，其中 $\widehat{M}^{(l)} = \frac{1}{N}\sum_{i=1}^{N}\left\{m(1, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_1^{(l)}) - m(0, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_0^{(l)})\right\}$且

$$\widehat{\delta}^{(l)} = \frac{1}{n_l}\sum_{i=1}^{n_l}\omega_l(X_i^{(l)}; \widehat{\boldsymbol{\eta}}^{(l)})\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1}\mathbf{1}(A_i^{(l)} = a)}{\pi_l(a, X_i^{(l)}; \widehat{\boldsymbol{\alpha}}^{(l)})}\{Y_i^{(l)} - m(A_i^{(l)}, X_i^{(l)}; \widehat{\boldsymbol{\gamma}}_a^{(l)})\}\right\}.$$

已证明，在相应的正则条件下，当结局模型正确设定，或密度比模型和倾向得分模型均正确设定时，双重稳健估计量 $\widehat{\theta}^{(l)}$ 渐近正态，其 $n_l^{1/2}(\widehat{\theta}^{(l)} - \theta^{(l)}) \xrightarrow{d} N(0, V^{(l)})$。我们使用影响函数展开，将方差 $V^{(l)}$ 估计为 $\widehat{V}^{(l)}$。在补充材料的附录A.3中，我们针对一组特定的模型选择给出了用于计算 $\widehat{V}^{(l)}$ 的影响函数。

由于 $\theta^{(l)}$是一维的，我们只需估计站点之间的差异度：$\mathcal{L}_{l,k} = \theta^{(l)} - \theta^{(k)}$，其标准误可估计为 $\text{标准误}(\widehat{\mathcal{L}}_{l,k}) = \sqrt{\widehat{V}^{(l)} + \widehat{V}^{(k)}}$，对于 $1 \le l, k \le L$。如注释1所述，RIFL估计量可以简化，并基于$\{\widehat{\mathcal{L}}_{l,k}, \text{标准误}(\widehat{\mathcal{L}}_{l,k}))\ 1 \le l < k \le L\}$ 以及 $\{(\widehat{\theta}^{(l)}, \widehat{V}^{(l)})\ l = 1, ..., L\}$ 构造。

6. 模拟研究

我们通过第 5.1、5.2 和 5.3 节讨论的三个推断问题展示 RIFL。针对这三个问题实现 RIFL 的代码可在 https://github.com/ celehs/RIFL 获取。我们还实现了另外三种估计量以供比较：基于中位数估计量的置信区间、m-out-of-n 自助法（MNB）置信区间，以及式（10）中的最大团投票（VMC）估计量。对于每种情形，我们进行 500 次模拟，以考察 95%置信区间的经验覆盖率和平均长度。我们考虑主导集合与其余站点之间的 5 种不同分离程度，以考察分离程度对不同方法性能的影响。为简单起见，我们将所有站点的样本量设为相同， $n_l = n$ 对于 $1 \le l \le L$，并考虑 $n = 500$、1000 和 2000年。全文中，我们令 $L = 10$ 并重点考察 $|\mathcal{V}(\theta^*)| = 6$ 的情形。

我们将中位数估计量构造为 $\{\widehat{\beta}^{(l)}\}_{1 \le l \le L}$ 的中位数，并通过参数自助法估计其标准误。对于 m-out-of-n 自助法（MNB），我们计算

^原始数据（记为 β^*）的点估计量，采用 VMC。对于每个站点，我们从 n 观测值中有放回地抽取 m 个观测值，并重复该过程 500 次以构造置信区间。根据 Bickel 和 Sakov 的研究，[2008]，子样本量 m 应相对于 n较小，因此我们设定 $m = n^{\upsilon}$，其中$\upsilon = 0.8$。对于 $1 \leq j \leq 500$，我们将 VMC 应用于第 j 个重采样数据，以计算点估计量$\widehat{\beta}^{m,\ j}$。我们定义 $L_n(t) = \frac{1}{500}\sum_{j=1}^{500} \mathbf{1}\{\sqrt{m}(\widehat{\beta}^{m,\ j} - \widehat{\beta}^*) \leq t\}$，其中$\mathbf{1}$ 表示指示函数。MNB 置信区间为

$$\left[\widehat{\beta}^* - \frac{\hat{t}_{1-\alpha/2}}{\sqrt{n}}, \widehat{\beta}^* - \frac{\hat{t}_{\alpha/2}}{\sqrt{n}}\right], \tag{38}$$

其中 $\hat{t}_{\alpha/2}$ and ^ $t_{1-\alpha/2}$是满足 $L_n(t) \geq \alpha/2$ 且 $L_n(t) \geq 1-\alpha/2$的最小 t，分别 .

此外，我们将预言偏差感知（OBA）置信区间作为基准，以^公平比较能够考虑选择变异性的方法。对于在（9）中定义的估计量 β^*，我们假设（$\widehat{\beta}^* - \beta^*$）/SE($\widehat{\beta}^*$) $\xrightarrow{d} N(b,\ 1)$，其中 SE($\widehat{\beta}^*$) 表示 β^* 的标准误^，b 表示相应的偏差。按照 Armstrong 等人 [2020]，在（7）中的做法，我们利用 $|\mathbf{E}\beta^* - \beta^*|$ 的^预言知识，并构造预言偏差感知置信区间如下：

$$\left(\widehat{\beta}^* - \chi, \widehat{\beta}^* + \chi\right) \quad \text{with} \quad \chi = \widehat{\mathrm{SE}}(\widehat{\beta}^*) \cdot \sqrt{\mathrm{cv}_\alpha\left(|\mathbf{E}\widehat{\beta}^* - \beta^*|^2/\widehat{\mathrm{SE}}^2(\widehat{\beta}^*)\right)}, \tag{39}$$

其中 $\mathrm{cv}_\alpha(B^2)$ 是自由度为 1、非中心参数为 B^2. 的 χ^2 分布的 $1-\alpha$ 分位数。我们进行了 500 次模拟，现说明一种预言方法，用于计算第 j 次模拟的^ ^估计标准误 SE(β^*)，其中 $1 \leq j \leq 500.$。对于第 j 次模拟，若 $1 \leq j \leq 500$，我们得到点估计量 $\widehat{\beta}^{*,\ j}$，其定义见（9），同时得到 SE $_j = \frac{1}{\sqrt{\sum_{t\in\hat{v}} 1/\hat{\sigma}_l^2}}$.。利用这 500 个估计量，我们计算 $\{\widehat{\beta}^{*,j}\}_{1\leq j\leq 500}$的经验标准误，并将其定义为 ESE。随后，我们计算 ESE 与 $\sum_{j=1}^{500}\mathrm{SE}_j/500$ 的比值，并将第 j 次模拟中的（39）式^SE($\widehat{\beta}^*$)定义为 ^SE($\widehat{\beta}^*$) $= \mathrm{ESE}\sum_{j=1}^{500}\mathrm{SE}_j/500 \cdot \mathrm{SE}^{j\cdot}$。我们认为，OBA 置信区间比（24）中假设已知多数群体的预言机置信区间更适合作为基准，因为它考虑了选择后误差。

6.1. 多源低维预测模拟

我们首先考察 RIFL 在第5.1小节所述的低维多源预测问题中的表现，该问题包含 $d = 10$ 个协变量和 $L = 10$ 个站点。协变量 $\{X_i^{(l)}\}_{1\leq i\leq n}$ 独立同分布地生成自均值为零、协方差为 $\Sigma \in \mathbb{R}^{d\times d}$ 的多元正态分布，其中 $\Sigma_{jk} = 0.6^{|j-l|}$ 对于 $1 \leq j,k \leq d$。对于满足 $1 \leq l \leq L$ 的第 l 个站点，我们生成二元结果 $Y(l)_i$，使其满足 P($Y(_{j}^{l} = 1|X_i^{(l)})$) $= 1/[1+\exp(-\mu_l - [X_i^{(l)}]^\top\theta^{(l)})]$,，其中 $1 \leq i \leq n_l$，站点特定截距为 μ_l。我们将 $(\mu_1, \ldots, \mu_{10})$ 设为 (0.05, −0.05, 0.1, −0.1, 0.05, −0.05, 0.1, −0.1, 0, 0)。对于主导集合中的站点，我们将 $\theta^{(l)} = \theta^* =$ 设为 (0.5, 0.5, 0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0, 0)，其中 $l \in \{1, \ldots, 6\}$。对于 $\theta^{(7)}, \theta^{(8)}, \theta^{(9)}$ 和 $\theta^{(10)}$，它们最后 5 个系数与 θ^* 相同，但前 5 个系数分别改为 $0.5 - 0.3a$, $0.5 - 0.2a$, $0.5 - 0.1a$ 和 $0.5 + 0.1a$，其中 $a \in \{1, 2, 3, 4, 5\}$ 控制多数站点与非多数站点之间的分离程度。定义 $\beta^{(l)} = \theta_1^{(l)}$，我们的推断目标为 $\beta^* = \theta_1^*$。站点特定估计量 $\{\widehat{\theta}^{(l)}\}_{1\leq l\leq L}$ 通过逻辑回归获得。在图2中，我们展示了基于各种方法进行 500 次模拟重复后，95%置信区间的经验覆盖率和平均长度。

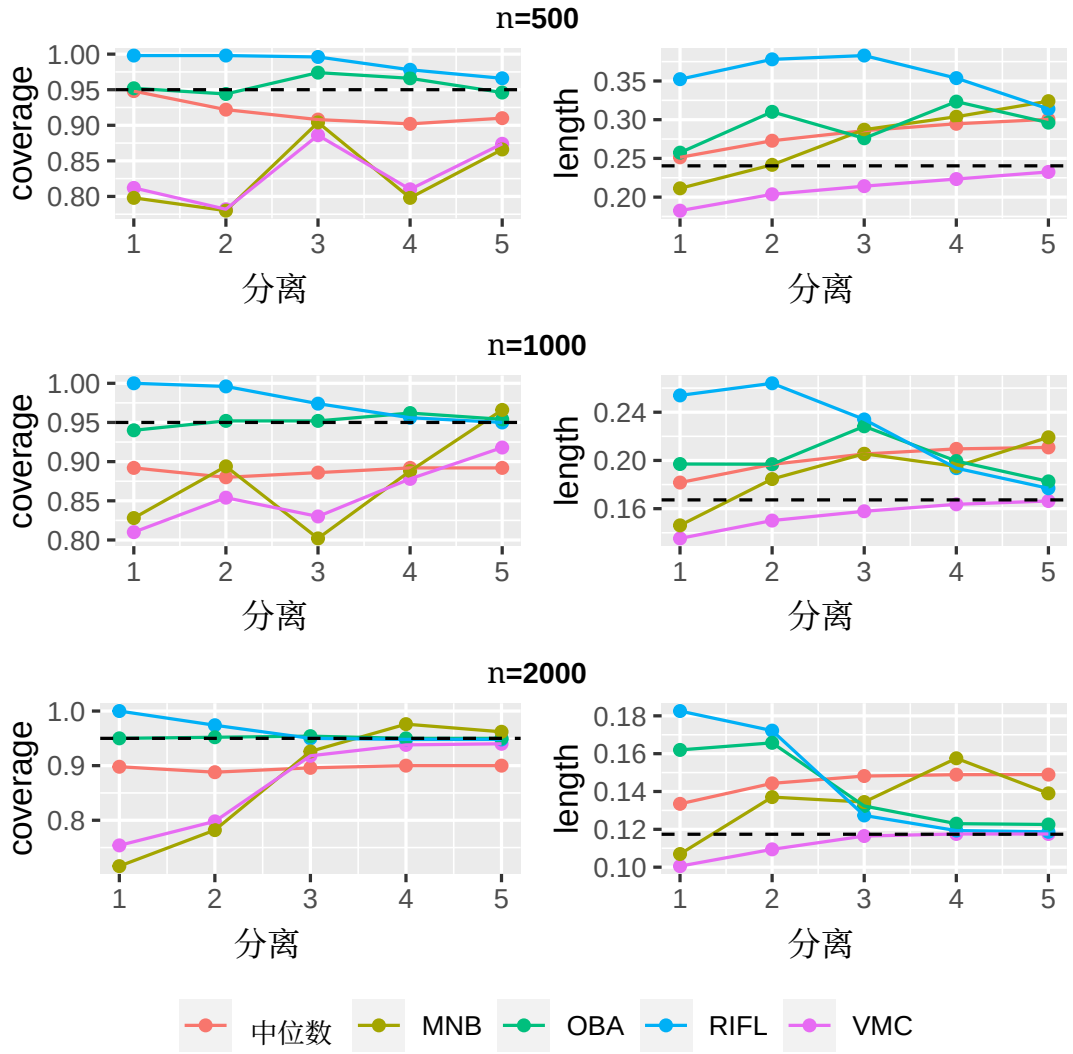

图 2：低维预测：对于 $\beta^* = \theta_1^*$，在有 6 个多数站点且分离程度各异的情况下，95%置信区间的覆盖率和长度；分离程度从 1（最难检测）到 5（最易检测）。“median”表示基于中位数估计量的置信区间，“MNB”表示式 (38) 中的 m-out-of-n 自助法置信区间，“VMC”表示式 (10) 中的最大团投票估计量及其相应的置信区间，“OBA”表示式 (39) 中的预言机偏差感知置信区间，“RIFL”表示式 (19) 中我们提出的置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

如图2所示，在所有分离程度下，RIFL置信区间均达到名义覆盖率；当分离程度较高时，其平均长度几乎与 OBA 一样短。当分离程度较低或中等时，VMC估计量的置信区间覆盖率低于名义覆盖率。只有在分离程度较高且样本量较大时，VMC估计量的覆盖率才接近名义覆盖率。这是由于在区分多数群体与

其余站点。如第 2.4 节所述，基于中位数估计量的置信区间由于其偏差，覆盖率低于名义水平。在这种情况下，由于非多数站点的参数值落在 β^* 的两侧，偏差适中，覆盖率通常仅下降至 90%。基于 m-out-of-n 自助法的置信区间在分离程度较低或中等时往往覆盖不足。这一观察结果与 Andrews [2000], Guo 等人 [2021a] 的讨论一致；在非正则情形下，MNB 可能无法提供有效推断。回顾一下，我们选择了 $m = n^{\upsilon}$ with $\upsilon = 0.8$。在补充材料的附录 C 中，我们展示了基于 υ 的各种选择所得的 MNB 结果。我们的结果表明，对于我们所考察的 υ 值范围，当分离程度较低时，MNB 方法往往覆盖不足。

6.2. 多源高维预测模拟

接下来，我们考察 RIFL 在第5.2节所述高维多源预测问题中的表现，该问题具有连续结局和 $d = 500$ 协变量。对于 $1 \le l \le L$，协变量 $\{X_i\}_{1 \le i \le n_l}$ 由均值为 0、协方差为 $\Sigma/2$ 的多元正态分布独立同分布地生成，其中 $\Sigma_{j,k} = 0.6^{|j-k|}$ 对于 $1 \le j, k \le d$。对于第 l 个来源站点，我们根据以下模型 $Y(_{i}{}^{l} = \mu_l + [X_i^{(l)}]^\top \theta^{(l)} + \varepsilon_i^{(l)})$ 生成结局 $Y(i)$，其中 $1 \le i \le n_l$，且 $\varepsilon_i^{(l)} \sim N(0, 1)$。我们令 $\mu_l = 0.05$ 对于 $l \in \{1, 2, 3, 7, 9\}$ 成立，否则为 0，并将 $\theta^{(l)} = \theta^*$ 设为 $l \in \{1, \ldots, 6\}$，其中 $\theta_j^* = (0.1j - 0.6)1(1 \le j \le 11)$，以表示稀疏信号和变化的信号强度。对于非多数站点，我们令 $\theta_j^{(7)} = \theta_j^{(8)} = \theta_j^* + 0.2 + 0.05a$，并令 $\theta_j^{(9)} = \theta_j^{(10)} = \theta_j^* + 0.15 + 0.05a$ 对于$6 \le j \le 11$成立。我们定义 $\beta^{(l)} = \theta_{11}^{(l)}$，并在下文中重点对 $\beta^* = \theta_{11}^*$ 进行推断。关于 θ_8^*的其他结果见补充材料中的附录 C。我们改变 a 在 $\{1, 2, 3, 4, 5\}$中的取值，以表示多数站点与非多数站点之间不同程度的分离。由于计算成本较高，我们不实施 m-out-of-n 自助法。

在图3中，我们展示了不同方法所得置信区间的经验覆盖率和平均长度。基于中位数估计量的置信区间覆盖率甚至低于第6.1节中的前一个示例；由于非多数站点的参数值存在方向相同的偏差，因此偏差更为严重。更具体地说，当有6个多数站点时，中位数估计量对应于多数站点中最大的站点特定估计量与第二大的站点特定估计量的平均值；更多讨论请参见第 2.4 节。

当 $n = 500$ 时，VMC置信区间的覆盖率始终偏低。这是因为在样本量有限的情况下，我们没有足够的统计功效来检测所考虑的最大分离程度（对应于 $a = 5$）。因此，一些非多数站点被错误地纳入估计的主导集合中。当样本量增加到 $n = 1000$ 和 2000 年时，分离程度足够大时，VMC置信区间可达到名义覆盖率，但分离程度较低时仍会出现覆盖不足。OBA区间和RIFL置信区间在所有设置下均达到预期覆盖率。然而，当 $n = 500$ 时，随着分离程度增大，OBA区间会变长。正如我们先前所讨论的，由于样本量有限，无论分离程度如何，一些非多数站点都会被纳入估计的主导集合。事实上，由这些错误选入的站点导致的VMC估计量偏差会随着分离程度增大而增大，因为非多数站点与多数站点之间的距离更远。OBA置信区间会考虑这一不断增大的偏差，因此会随着分离程度增大而变长。此外，当 $n = 1000$ 时，在分离程度较低至中等的情况下，OBA区间远长于RIFL区间。这同样是因为OBA区间会校正VMC估计量的偏差，而该偏差可能与VMC估计量的经验标准误一样大。在所有设置下，RIFL置信区间的长度与OBA区间相当，而且通常更短。

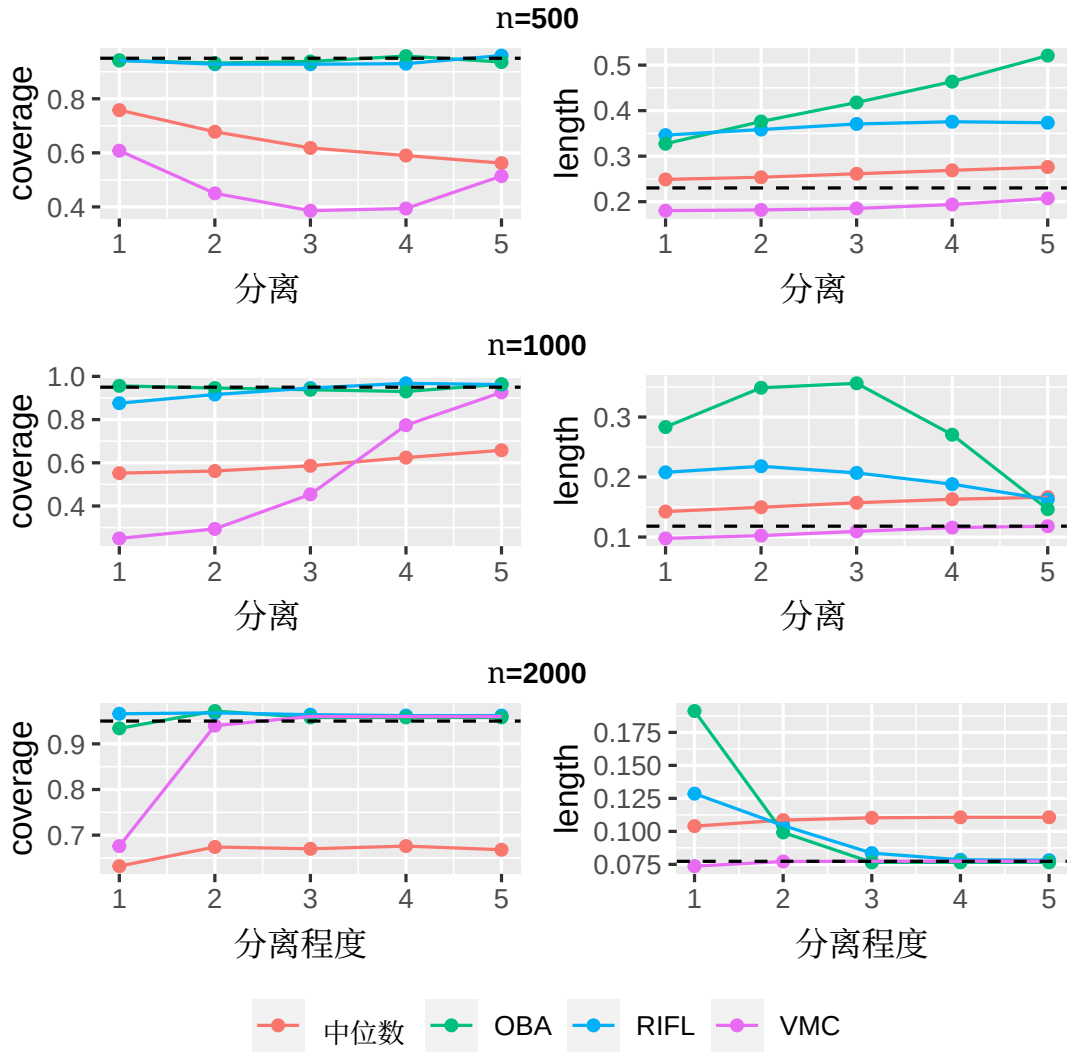

图 3： $\beta^* = \theta^*_{11}$ 的高维预测：在有 6 个多数站点且分离程度各异的情况下，95%置信区间的覆盖率和长度，其中 1 表示最低（最难检测），5 表示最高（最易检测）。“中位数”表示基于中位数估计量的置信区间，“VMC”表示最大团投票估计量及其在（10）中的相应置信区间，“OBA”表示（39）中的预言机偏差感知置信区间，“RIFL”表示（19）中所提出的置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应于 0.95 的名义覆盖率；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

6.3. 多源因果模拟

我们评估了RIFL在多源因果建模问题中的表现；该问题关注的是对目标总体的平均处理效应进行推断，如第5.3小节所述。我们考虑一个10维协变量向量，即 $p = 10$。对于 $l \in \{1, \ldots, L\}$，协变量 $\{X_i^{(l)}\}_{1 \le i \le n_l}$ 按照多元正态分布独立同分布地生成，其分布为

均值 $\mu_X^{(l)}$ 和协方差矩阵 Σ，其中 $\Sigma_{jk} = 0.6^{|j-k|}$ 适用于 $1 \le j$、$k \le p$。对于 $l \in \{1,2,3,7,9\}$，我们将 $\mu_X^{(l)}$ 设为 0；对于 $l \in \{4,5,6,8,10\}$，我们将 $\mu_X^{(l)}$ 设为 (0.5, 0.5, 0, 0, . . . , 0)。对于 $1 \le i \le n_l$，给定 $X_i^{(l)}$，治疗分配根据条件分布$A_i^{(l)}|X_i^{(l)} \sim$ 伯努利分布{逻辑函数($\alpha_1^{(l)} X_{i1}^{(l)} + \alpha_2^{(l)} X_{i2}^{(l)} + \alpha_{12}^{(l)} X_{i1}^{(l)} X_{i2}^{(l)}$)}生成，其中 $\alpha_1^{(l)} = 0.5$、$\alpha_2^{(l)} = -0.5$ 和$\alpha_{12}^{(l)} = 0.1$。第 l 个站点的结局根据 $Y_i^{(l)} = \mu_l + [X_i^{(l)}]^\top \zeta^{(l)} + \beta^{(l)} A_i^{(l)} + \varepsilon_i^{(l)}$ 生成，其中 $\varepsilon_i^{(l)} \sim N$对于 $1 \le i \le n_l$.服从 (0, 1)。这里，所有站点的回归系数 $\zeta^{(l)}$ 均设为 (0.5, 0.5, 0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0, 0)，站点特定截距 $(\mu_1, \mu_2, \ldots, \mu_{10})$ 设为 (0.05, −0.05, 0.1, −0.1, 0.05, −0.05, 0.1, −0.1, 0, 0)。系数 $\beta^{(l)}$ 衡量治疗效应。对于多数站点，即 $l \in \{1, \ldots, 6\}$, $\beta^{(l)} = -1$。我们设定 $\beta^{(7)} = \beta^{(8)} = -1 - 0.2a$ 和 $\beta^{(9)} = \beta^{(10)} = -1 - 0.1a$，其中 $a \in \{1,2,3,4,5\}$ 控制分离程度。目标总体中的协变量从均值为 0、协方差矩阵为 Σ 的多元正态分布中生成。在评估平均处理效应时，我们从目标总体生成了 $N = 10,000$ 个 X 实现值。

为估计结局回归，我们在每个站点的每个治疗组内，拟合一个包含各协变量主效应的线性模型。倾向评分通过逻辑回归模型估计，该模型仅在每个站点内包含 $X_{\cdot 1}^{(l)}$、和 $X_{\cdot 2}^{(l)}$、的主效应。最后，我们假设密度比满足以下模型： $\omega_l(X_i^{(l)}; \boldsymbol{\eta}^{(l)}) = \exp([\boldsymbol{\eta}^{(l)}]^\top \widetilde{X}_i^{(l)})$，其中 $\widetilde{X}_i^{(l)} = (1, (X_i^{(l)})^\top)^\top$。我们通过求解估计方程 $\sum_{i=1}^n \exp([\boldsymbol{\eta}^{(l)}]^\top \widetilde{X}_i^{(l)})\, \widetilde{X}_i^{(l)}/n = \sum_{j=1}^N \widetilde{X}_j^{\mathcal{T}}/N$来估计$\boldsymbol{\eta}^{(l)}$，其中 $\widetilde{X}_j^{\mathcal{T}} = (1, (X_j^{\mathcal{T}})^\top)^\top$，且 $X_j^{\mathcal{T}}$表示目标数据集中第 j个观测值，求解对象为 $j \in \{1, \ldots, N\}$。值得注意的是，尽管倾向得分模型设定有误，只要结局回归模型设定正确，双重稳健性仍能保证所得估计量相合且渐近正态。

基于 500 次模拟重复验证，通过各种方法得到的置信区间（CI）的经验覆盖率和平均长度图见图4，设置为 6 个多数站点。我们观察到与第6.1节低维预测示例中相似的模式。具体而言，随着分离程度和样本量增加，VMC 置信区间的覆盖率趋近于名义水平，但当分离程度较小至中等时，覆盖率通常低于名义水平。分离程度不够大时，MNB 置信区间的覆盖率也不足。RIFL 和 OBA 在所有设置下均达到名义覆盖率，且其长度通常相近。在这种情况下，基于中位数估计量的置信区间覆盖率较低，因为非多数站点的平均处理效应均小于多数站点的平均处理效应，因此中位数估计量存在严重偏差。

6.4. 对不同调优参数选择的稳健性

我们通过实证评估 RIFL置信区间的覆盖率和精度对调优参数 $\rho(M)$ 及重抽样规模 M. 选择的敏感性。我们继续研究第6.1节中的低维多源预测问题，其中多数规则由 6 个多数站点满足，且 $n_l = n = 1000$，适用于 $1 \le l \le 10$。此外，我们设定分离程度为 $a = 3$。如图5所示，就经验覆盖率和平均长度而言，RIFL置信区间随重抽样规模 M （500、1,000 或 5,000）的不同选择仅略有变化。正如第3.3节所述，我们选择最小的 $\rho(M)$，使得超过一定比例（prop）的重抽样集合满足多数规则。我们通过将比例（prop）

从 5% 到 50% 按每次增加 5% 进行调整，以检验对 $\rho(M)$ 选择的敏感性。在图5中，我们观察到，采用不同 $\rho(M)$ 选择的 RIFL置信区间均达到了预期覆盖率。RIFL置信区间的平均长度随比例增加。

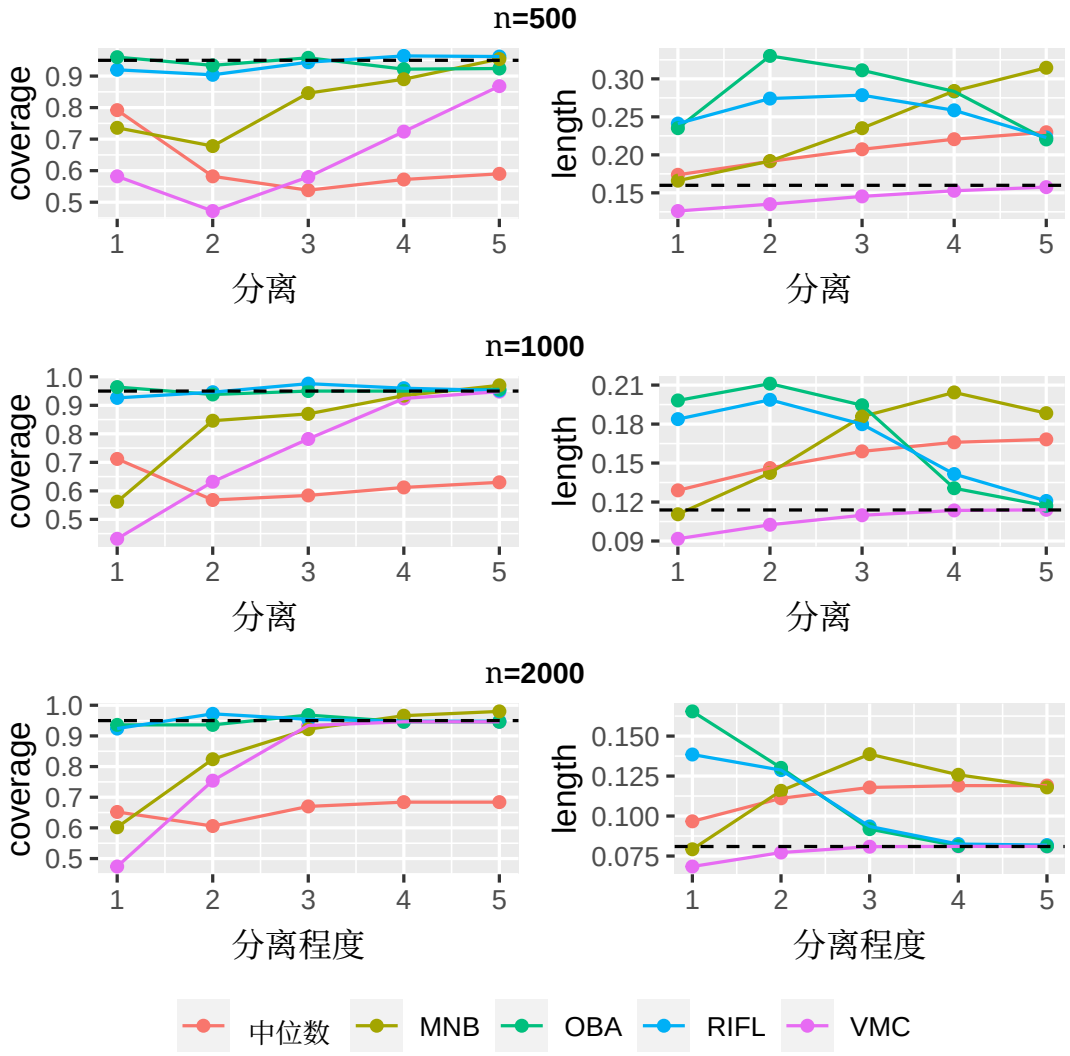

图4：因果推断：在有6个多数站点且分离程度不同的情况下，多数站点目标平均处理效应的95%置信区间覆盖率和长度，其中1表示最低分离程度（最难检测），5表示最高分离程度（最易检测）。“中位数”表示基于中位数估计量的置信区间；“MNB”表示式（38）中的m-out-of-n自助法置信区间；“VMC”表示最大团投票估计量及其在式（10）中的相应置信区间；“OBA”表示式（39）中的偏差感知预言置信区间；“RIFL”表示我们在式（19）中提出的置信区间。结果基于500次模拟重复。左侧面板中的虚线对应名义覆盖率0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

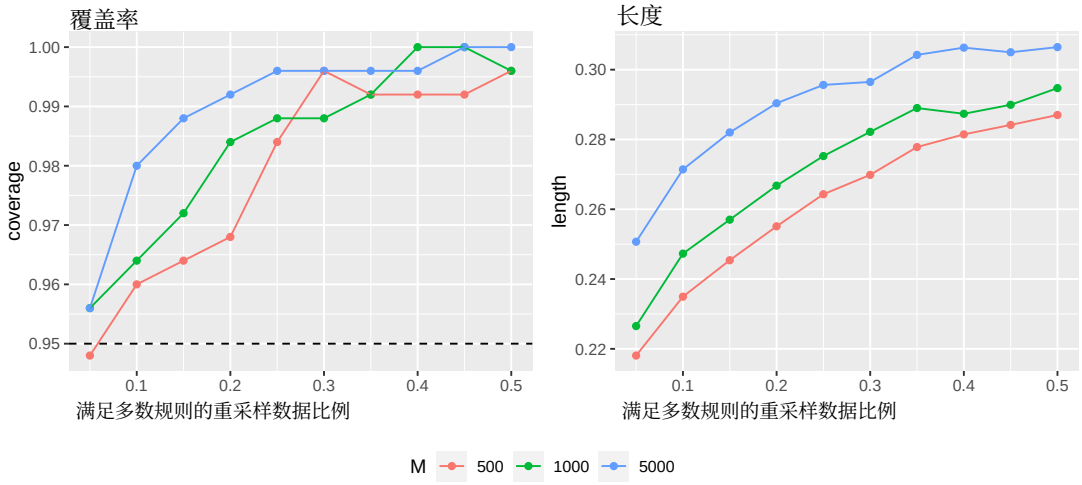

图 5：RIFL置信区间的覆盖率和平均长度对不同 M 和 $\rho(M)$ 的敏感性。结果基于 200 次模拟重复。横轴表示 prop 的取值，其中调优参数 $\rho(M)$ 选为满足以下条件的最小值：超过 prop 的重抽样差异度量会产生满足多数规则的最大团集。

7. RIFL：基于电子健康记录的 COVID-19死亡预测研究综合分析

严重急性呼吸综合征冠状病毒2型（SARS-CoV-2）已导致数百万例 COVID-19 感染和死亡。尽管大多数感染者仅出现轻度病毒性肺炎，但仍有一部分人会发展为重症。对于部分人群而言，COVID-19 仍是一种致命疾病，包括老年人和免疫系统受损者 [Wu 和 McGoogan，2020，Goyal 等人，2020，Wynants 等人， 2020]。识别死亡风险较高的患者有助于挽救生命，并改善资源匮乏的卫生系统中的资源分配。

为了更好地了解 COVID-19 住院患者的死亡风险，我们使用来自四个国家的 $L = 16$ 医疗中心汇总的数据实施 RIFL 算法。这些数据代表了 275 家医院，属于多机构基于电子健康记录的 COVID-19 临床特征联盟（4CE） [Brat 等人， 2020]。纳入研究的患者必须聚合酶链式反应（PCR）检测呈 SARS-CoV-2 阳性；在 2020年3月1日至2021年1月31日期间，因 PCR 检测阳性而住院；且住院时间须在首次 PCR 检测阳性日期前不超过七天、之后不超过 14 天。入院当天死亡的患者被排除。在 RIFL 分析中，我们聚焦于可获得汇总层面数据（包括死亡率回归模型的协方差矩阵）的站点子集，共计分析 42,655 名患者。由于患者隐私限制，个体层面数据仍保留在各机构的防火墙内；与此同时，已为各类研究项目生成汇总层面数据 [，例如 Weber 等人， 2022]。

我们重点关注在全部 16 个站点测量的基线风险因素。这些基线风险因素包括年龄组（18-25、26-49、50-69、70-79、 80+）、性别、入院前 Charlson 合并症评分，以及入院时的九项实验室检测值：白蛋白、天门冬氨酸氨基转移酶（天冬氨酸氨基转移酶，

天门冬氨酸氨基转移酶与丙氨酸氨基转移酶比值（AST/ALT比值）、胆红素、肌酐、C反应蛋白（C反应蛋白）、淋巴细胞计数、中性粒细胞计数和白细胞（WBC）计数。实验室检测数据的缺失比例较低（$< 30\%$），缺失值采用链式方程多重插补法进行插补，并对五组插补数据集的结果取平均。

我们的目标是识别并估计一个当前普遍使用的死亡风险预测模型，以更好地理解基线风险因素对死亡率的影响。具体而言，对于每个医疗中心 l、$1 \leq l \leq 16$，我们令 $\theta^{(l)} \in \mathbb{R}^{15}$ 表示与15个风险因素相关的对数风险比向量，即多变量Cox模型中的回归系数向量。我们对与每个风险因素相关的对数风险比进行统计推断，并令 $\beta^{(l)} = \theta_j^{(l)}$，同时令 j 从1变动到15。我们采用自适应LASSO惩罚拟合Cox模型，以获得 $\widehat{\theta}^{(l)}$，即15个基线风险因素对住院后30天内全因死亡率的估计对数风险比。由于样本量相对于 $p = 15$而言较大，我们预计自适应LASSO估计量的神谕性质能够成立，因此可以依靠这些局部估计量的渐近正态性，以及其方差的一致估计量 [Zou、 2006]。

我们在图 6 中同时实现 RIFL 和 VMC，为每个基线风险因素生成 95%置信区间。对于 RIFL，我们设置 $M = 500$，并根据第3.3节选择 $\rho(M)$ 的值。也就是说，我们从一个较小的值（例如 $1/12$）开始，并将 $\rho(M)$ 设为小于 1 的最小值，使得超过 10% 的重抽样不相似性度量会生成满足多数规则的最大团集。可推广性度量 $\{\widehat{p}_l\}_{1 \leq l \leq 16}$，定义为 $\widehat{p}_l = \sum_{m \in \mathcal{M}} \mathbf{1}(l \in \widetilde{\mathcal{V}}^{[m]})/|\mathcal{M}|$，表示医疗中心 l 被纳入多数群体的比例。请注意，对于每个风险因素，我们都可以计算 16 个站点的一组可推广性度量。补充材料中的图 C7 展示了完整可推广性度量的可视化结果。

如图 6 所示，RIFL 区间与 VMC 区间大多具有可比性，尽管对于某些风险因素，RIFL置信区间明显长于 VMC置信区间。然而，正如我们在模拟研究中所展示的那样，由于站点选择错误，VMC置信区间可能存在覆盖不足。事实上，当我们聚焦于某一特定风险因素时，一些站点的可推广性评分较低。例如，当我们聚焦于18至25岁年龄组这一变量时，站点 4 的可推广性度量为 0.020；当我们聚焦于年龄 > 80 这一变量时，站点 2 的可推广性度量为 0.016。这表明，在研究相应风险因素的影响时，这些站点与其他站点的匹配程度并不高。然而，VMC 将这些站点纳入了估计的主导集合，由此得到的置信区间可能会产生误导。总体而言，在全部 15 个风险因素中，除站点 2 和 4 外，大多数站点都具有较高的可推广性度量，如补充材料中的图 C7 所示。

我们的结果表明，对于聚合酶链式反应检测 COVID-19 呈阳性并住院的患者，年龄较大、男性、Charlson 合并症评分较高、血清白蛋白水平较低，以及肌酐、C反应蛋白和天冬氨酸氨基转移酶水平较高，均与较高的死亡风险相关。能够预测死亡率的实验室检测指标涵盖一般健康状况（如血清白蛋白）、肾功能（如肌酐）、肝功能（如天冬氨酸氨基转移酶）以及急性炎症反应（如C反应蛋白）。这些结果与相关研究中利用实验室检测结果预测 COVID-19死亡率的既往发现相合 [Weber 等人, 2022]。近期文献表明，检测血清白蛋白有助于识别更可能发展为重症的 COVID-19 感染患者，且血清白蛋白可作为疾病进展的有效标志物 [Turcato 等人，2022]。我们的结果进一步印证了这一发现。除基线人口统计学和合并症信息外，使用常规采集的实验室检测结果，也有助于开发临床实用的预测工具，用于评估 COVID-19 患者住院后的死亡风险。随后，可优先对死亡风险较高的患者进行更密切的监测，并在医师认为适当时采取

更积极的干预措施。

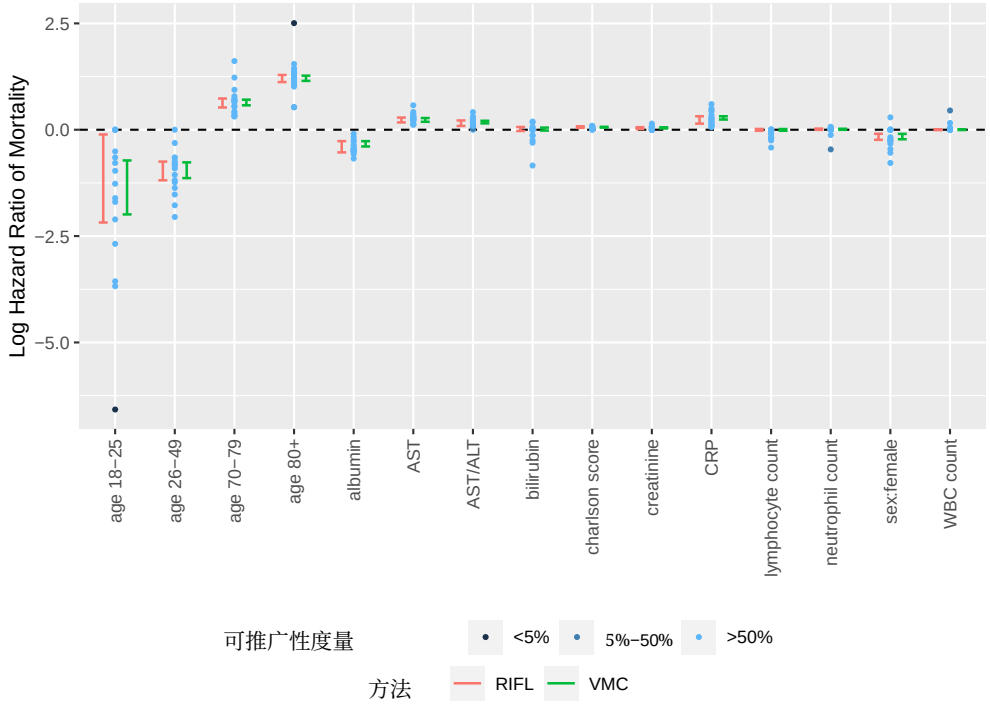

图 6：RIFL 和 VMC 对 COVID-19 患者住院后 14 天内死亡率对数风险比的 95%置信区间，涵盖 15 个基线风险因素中的每一个。圆点表示各个站点的点估计，圆点颜色越深，表示该站点的可推广性度量越低。

8. 结论与讨论

RIFL置信区间能够抵御将非多数站点与多数站点区分开时产生的误差，并保证对主导模型的一致有效覆盖。当多数站点与其他站点明确分离时，RIFL置信区间的表现与假设已知多数群体的预言机置信区间相似。

多数规则是我们提出的方法的一项关键假设，对多数规则进行经验评估对于验证我们提出的程序至关重要。我们提出的RIFL方法或可对多数规则进行启发式评估。假设在 (16) 中定义的索引集 $\mathcal{M}$ 的基数低于 $10\% \cdot M$，即使 ρ 接近 1 也是如此。在这种情况下，我们认为多数规则可能不成立，因为难以利用大多数重采样数据验证多数规则。作为对多数规则的放宽，我们可以将 (1) 中 $\mathcal{V}(\theta)$ 定义中的严格相等放宽为近似相等，

$$\mathcal{V}_\kappa(\theta) := \{1 \le l \le L : \|\theta^{(l)} - \theta\|_2 \le \kappa\}, \quad \text{for some} \quad \kappa > 0. \tag{40}$$

当 κ 足够接近零时，我们的主要结果或许可以推广，使其在基于 $\mathcal{V}_\kappa(\theta)$定义的多数规则（假设 1）下仍然成立。

在实践中，领域专家可能认为，总站点中超过 50%（例如 80%）的站点具有相似性。我们提出的 RIFL 可以直接扩展，以纳入额外的先验信息。例如，我们可以考虑 80%规则；也就是说，领域专家预期超过 80% 的站点具有相同的目标参数。我们可以通过将（16）和（17）中的 $L/2$ 替换为 $80\% \cdot L$. 来修改构造。重要的是，只要主导集合中确实有超过 80% 的站点，利用多数规则或 80%规则的 RIFL 方法都能保证对主导模型参数的置信区间具有一致有效的覆盖率。然而，由于额外的先验信息，利用 80%规则的 RIFL 得到的区间长度会更短；补充材料 C.1 节展示了我们对不同先验知识下 RIFL 置信区间覆盖率和长度的比较。也可以考虑其他策略来选择 ρ，以便基于 $\{\widehat{p}_l\}_{1\le l\le L}$进一步进行主导集合大小的数据自适应评估。如何从理论上证明自适应选择 $\rho(M)$的方法仍有待进一步研究。

当多数规则或其宽松版本不成立时，我们讨论稳健信息聚合的两个可能方向。首先，可以将目标预测模型定义为最多站点共同采用的模型。如果我们根据相似性将 $\{\theta^{(l)}\}_{1\le l\le L}$ 划分为若干子组，那么规模最大的组自然可以定义一个目标模型，即使该组所包含的站点不超过 $L/2$个。其次，我们可以在（40）中定义集合 $\mathcal{V}_\kappa(\theta)$，并令 κ相对较大；也就是说， $\mathcal{V}_\kappa(\theta)$包含所有参数与 θ相似但可能有所不同的站点。我们仍可以通过定义群体分布鲁棒模型（如 Meinshausen 和 B¨ uhlmann [2015], Hu 等人 [2018], Sagawa 等人 [2019], Guo [2020]所述），汇集属于 $\mathcal{V}_\kappa(\theta)$的信息。这两种情形下都存在后选择问题。考虑识别有用站点时的不确定性，可以说至关重要。我们提出的方法有望用于处理站点选择问题，并对这些目标模型进行推断；这留待未来研究。

致谢

本研究部分得到了美国国家科学基金会（NSF）资助的 DMS 2015373 项目，以及美国国立卫生研究院（NIH）资助的 R01GM140463、R01HL089778 和 R01LM013614 项目的支持。

参考文献

Donald WK Andrews. 当参数位于参数空间边界上时自助法的不相合性。《计量经济学》，第399–405页，2000年。Martin Arjovsky、L´eon Bottou、Ishaan Gulrajani 和 David Lopez-Paz。风险不变性最小化。*arXiv*预印本 *arXiv:1907.02893*，2019年。Timothy B Armstrong、Michal Koles´ar 和 Soonwoo Kwon。正则化回归模型中的偏差感知推断。*arXiv* 预印本 *arXiv:2012.14823*，2020年。Heejung Bang 和 James M Robins。缺失数据与因果推断模型中的双重稳健估计。《生物统计学》，61(4):962–973，2005年。Alexandre Belloni、Victor Chernozhukov 和 Christian Hansen。高维控制变量中进行选择后的处理效应推断。《经济研究评论》，81(2):608–650，2014年。

Richard Berk、Lawrence Brown、Andreas Buja、Kai Zhang 和 Linda Zhao。有效的后选择推断。《统计年鉴》，41(2):802–837，2013年。Peter J Bickel 和 Anat Sakov。如何选择m取n自助法中的m，以及极值的置信界。《中华统计学志》，第967–985页，2008。Peter J Bickel、Ya'acov Ritov 和 Alexandre B Tsybakov。套索回归与Dantzig选择器的同步分析。《统计年鉴》，37(4):1705–1732，2009年。Jack Bowden、George Davey Smith、Philip C Haycock 和 Stephen Burgess。使用加权中位数估计量，在存在部分无效工具变量的孟德尔随机化中进行相合估计。《遗传流行病学》，40(4):304–314，2016年。Gabriel A Brat、Griffin M Weber、Nils Gehlenborg、Paul Avillach、Nathan P Palmer、Luca Chiovato、James Cimino、Lemuel R Waitman、Gilbert S Omenn、Alberto Malovini 等。国际电子健康记录衍生的COVID-19 临床病程概况：4CE联盟。《*NPJ*数字医学》，3(1):1–9，2020年。Peter B¨uhlmann 和 Nicolai Meinshausen。Magging：针对非同质大规模数据的极大极小聚合。《*IEEE*会刊》，104(1):126–135，2015年。Peter B¨uhlmann 和 Sara van de Geer。《高维数据统计：方法、理论与应用》。施普林格科学与商业媒体，2011年。Stephen Burgess、Dylan S Small 和 Simon G Thompson。孟德尔随机化工具变量估计量综述。《医学研究统计方法》，26(5):2333–2355，2017年。T Tony Cai 和 Zijian Guo。高维线性回归的置信区间：极小极大速率与自适应性。《统计年鉴》，45(2):615–646，2017年。T. Tony Cai 和 Zijian Guo。高维线性回归中解释方差的半监督推断及其应用。《英国皇家统计学会会刊：*B*辑（统计方法学）》，82(2):391–419，2020年。T Tony Cai、Zijian Guo 和 Rong Ma。二元结局高维广义线性模型的统计推断。《美国统计学会杂志》，第1–14页，2021a。Tianxi Cai、Molei Liu 和 Yin Xia。高维异质性数据的个体数据保护型整合回归分析。《美国统计学会杂志》，第1–15页，2021b。Tianxi Cai、T Tony Cai 和 Zijian Guo。高维模型中个体化治疗效应的最优统计推断。《英国皇家统计学会会刊：*B*辑（统计方法学）》，83(4):669–719，2021c。Randy Carraghan 和 Panos M Pardalos。最大团问题的精确算法。《运筹学快报》，9(6):375–382，1990年。Bibhas Chakraborty、Eric B Laber 和 Yingqi Zhao。使用自适应m取n自助法方案对最优动态治疗方案进行推断。《生物统计学》，69(3):714–723，2013年。

Xueying Chen 和 Min-ge Xie。《超大规模数据分析的分而治之方法》。《中华统计学志》，第1655–1684 页，2014年。Yixin Chen、Guozhu Dong、Jiawei Han、Jian Pei、Benjamin W Wah 和 Jianyong Wang。《采用无损压缩和聚合的回归立方体》。*IEEE* 知识与数据工程汇刊，18(12):1585–1599，2006年。Victor Chernozhukov、Christian Hansen 和 Martin Spindler。《含有众多控制变量和工具变量的线性模型中的后选择与后正则化推断》。2015年。John Chodera、Alpha A Lee、Nir London 和 Frank von Delft。《为大流行病药物发现开展众包》。《自然·化学》，12(7):581–581，2020年。Ruthie Davi、Nirosha Mahendraratnam、Arnaub Chatterjee、C Jill Dawson 和 Rachel Sherman。《利用外部对照为单臂临床试验提供信息》。《自然综述：药物发现》，19 (12):821–822，2020年。Rui Duan、Chongliang Luo、Martijn J Schuemie、Jiayi Tong、C Jason Liang、Howard H Chang、Mary Regina Boland、Jiang Bian、Hua Xu、John H Holmes 等。《从局部学习到全局：一种用于生存时间数据建模的高效分布式算法》。《美国医学信息学会杂志》，27(7):1028–1036，2020年。Parag Goyal、Justin J Choi、Laura C Pinheiro、Edward J Schenck、Ruijun Chen、Assem Jabri、Michael J Satlin、Thomas R Campion Jr、Musarrat Nahid、Joanna B Ringel 等。《纽约市COVID-19 患者的临床特征》。《新英格兰医学杂志》，382(24):2372– 2374，2020年。Zijian Guo。《采用抽样方法对异质性回归模型中的高维极大极小效应进行推断》。*arXiv* 预印本 *arXiv:2011.07568*，2020年。Zijian Guo。《无效工具变量下的因果推断：后选择问题及一种基于搜索与抽样的解决方案》。*arXiv* 预印本 *arXiv:2104.06911*，2021年。Zijian Guo、Hyunseung Kang、T Tony Cai 和 Dylan S Small。《采用带投票的两阶段硬阈值法，计算无效工具变量下因果效应的置信区间》。《英国皇家统计学会会刊：*B*辑（统计方法学）》，80(4):793–815，2018年。Zijian Guo、Prabrisha Rakshit、Daniel S Herman 和 Jinbo Chen。《高维逻辑回归中的病例概率推断》。《机器学习研究杂志》，22(1):11480– 11533，2021a。Zijian Guo、Claude Renaux、Peter B ¨uhlmann 和 Tony Cai。《高维组推断及其在层级检验中的应用》。《电子统计学杂志》，15(2):6633–6676，2021b。Larry Han、Angela Chen、Jason J Ong、Juliet Iwelunmor 和 Joseph D Tucker。《健康与健康研究中的众包：实用指南》。2018年。Larry Han、Jue Hou、Kelly Cho、Rui Duan 和 Tianxi Cai。《目标治疗效应的联邦自适应因果估计（FACE）》。*arXiv* 预印本 *arXiv:2112.09313*，2021年。Reid Hastie 和 Tatsuya Kameda。《群体决策中多数规则的稳健之美》。心理学评论，112(2):494，2005年。

Neil T Heffernan 和 Cristina Lindquist Heffernan。《ASSISTments 生态系统：构建一个让科学家和教师携手开展人类学习与教学微干预研究的平台》。《国际人工智能教育杂志》，24(4):470–497，2014年。Weihua Hu、Gang Niu、Issei Sato 和 Masashi Sugiyama。《分布鲁棒监督学习能否产生鲁棒分类器？》载于国际机器学习大会，第 2029– 2037 页。PMLR，2018年。Mahta Jahanshahi、Keith Gregg、Gillian Davis、Adora Ndu、Veronica Miller、Jerry Vockley、Cecile Ollivier、Tanja Franolic 和 Sharon Sakai。《外部对照在美国食品药品监督管理局监管决策中的应用》。《治疗创新与监管科学》，55(5):1019–1035，2021年。Adel Javanmard 和 Andrea Montanari。《高维回归的置信区间与假设检验》。《机器学习研究杂志》，15(1):2869–2909，2014年。Hyunseung Kang、Anru Zhang、T Tony Cai 和 Dylan S Small。《含有部分无效工具变量的工具变量估计及其在孟德尔随机化中的应用》。《美国统计学会杂志》，111(513):132–144，2016年。Joseph DY Kang 和 Joseph L Schafer。《揭开双重稳健性的神秘面纱：从不完整数据估计总体均值的替代策略比较》。《统计科学》，22 (4):523–539，2007年。Norbert L Kerr、R Scott Tindale 等。《群体表现与决策》。《心理学年度评论》，55(1):623–655，2004年。Kevin L Keys、Angel CY Mak、Marquitta J White、Walter L Eckalbar、Andrew W Dahl、Joel Mefford、Anna V Mikhaylova、Mar´ıa G Contreras、Jennifer R Elhawary、Celeste Eng 等。《基因表达预测模型的跨人群泛化能力》。《公共科学图书馆·遗传学》，16(8): e1008927，2020年。Peter Kraft、Eleftheria Zeggini 和 John PA Ioannidis。《全基因组关联研究中的重复验证》。《统计科学：数理统计学会评论期刊》，24(4): 561，2009年。Jason D Lee、Dennis L Sun、Yuekai Sun 和 Jonathan E Taylor。《精确的后选择推断及其在套索回归中的应用》。《统计年刊》，44(3):907–927，2016年。Jason D Lee、Qiang Liu、Yuekai Sun 和 Jonathan E Taylor。《通信高效稀疏回归》。《机器学习研究杂志》，18(1):115–144，2017年。Hannes Leeb 和 Benedikt M P¨otscher。《模型选择与推断：事实与虚构》。计量经济学理论，第 21–59 页，2005年。Jeffrey T Leek、Robert B Scharpf、H´ector Corrada Bravo、David Simcha、Benjamin Langmead、W Evan Johnson、Donald Geman、Keith Baggerly 和 Rafael A Irizarry。《应对批次效应对高通量数据造成的普遍且严重影响》。《自然综述：遗传学》，11(10): 733–739，2010年。Runze Li、Dennis KJ Lin 和 Bing Li。《海量数据集中的统计推断》。《商业与工业中的应用随机模型》，29(5):399–409，2013年。

Sai Li、T Tony Cai 和 Hongzhe Li。高维线性回归的迁移学习：预测、估计与极小极大最优性。*arXiv* 预印本 *arXiv:2006.10593*，2020年。Heng Lian 和 Zengyan Fan。超高维中去偏 l1 范数支持向量机的分而治之方法。《机器学习研究杂志》，18(1):6691–6716，2017年。Wodan Ling、Jiuyao Lu、Ni Zhao、Anju Lulla、Anna M Plantinga、Weijia Fu、Angela Zhang、Hongjiao Liu、Hoseung Song、Zhigang Li 等。通过条件分位数回归去除微生物组数据中的批次效应。《自然·通讯》，13(1):1–14，2022年。Molei Liu、Yin Xia、Kelly Cho 和 Tianxi Cai。在数据共享约束下对具有异质性的高维数据进行整合式多重检验。机器学习研究杂志，22:126–1，2021年。Subha Maity、Yuekai Sun 和 Moulinath Banerjee。异质性数据的荟萃分析：高维整合式稀疏回归。《机器学习研究杂志》，23(198):1–50，2022年。Raymond H Mak、Michael G Endres、Jin H Paik、Rinat A Sergeev、Hugo Aerts、Christopher L Williams、Karim R Lakhani 和 Eva C Guinan。利用群体创新开发基于人工智能的放射治疗靶向解决方案。《*JAMA*肿瘤学》，5(5):654–661，2019年。Nicolai Meinshausen 和 Peter B¨uhlmann。非同质大规模数据中的极大极小效应。《统计年鉴》，43(4):1801–1830，2015年。Yang Ning 和 Han Liu。稀疏高维模型假设检验与置信区域的一般理论。《统计年鉴》，45(1):158–195，2017年。Jonas Peters、Peter B¨uhlmann 和 Nicolai Meinshausen。利用不变预测进行因果推断：识别与置信区间。《英国皇家统计学会会刊：*B*辑（统计方法学）》，78(5):947–1012，2016年。Laila Rasmy、Yonghui Wu、Ningtao Wang、Xin Geng、W Jim Zheng、Fei Wang、Hulin Wu、Hua Xu 和 Degui Zhi。利用大型异质电子健康记录数据集研究基于循环神经网络的预测模型对心力衰竭发病风险的可推广性。《生物医学信息学杂志》，84:11–16，2018年。Dominik Rothenh¨ausler、Nicolai Meinshausen 和 Peter B¨uhlmann。非同质大规模数据中极大极小效应的置信区间。载于《高维数据统计分析》，第255–277页。施普林格，2016年。Shiori Sagawa、Pang Wei Koh、Tatsunori B Hashimoto 和 Percy Liang。针对群体偏移的分布鲁棒神经网络：正则化对最坏情况下可推广性的重要性。*arXiv* 预印本 *arXiv:1911.08731*，2019年。Ryan G Short、Dana Middleton、Nicholas T Befera、Raj Gondalia 和 Tina D Tailor。以患者为中心的放射学报告：利用在线众包评估基于网络的交互式放射学报告的有效性。《美国放射学会杂志》，14(11):1489–1497，2017年。Robert D Sorkin、Ryan West 和 Donald E Robinson。群体表现取决于多数规则。《心理科学》，9(6):456–463，1998年。

James Surowiecki。群体智慧。Anchor，2005年。Yebin Tao 和 Haoda Fu。针对目标总体的加权平均处理效应的双重稳健估计。《医学统计学》，38(3):315–325，2019年。Ye Tian 和 Yang Feng。高维广义线性模型下的迁移学习。《美国统计学会杂志》，(刚刚接收)：1–30，2022年。Jiayi Tong、Chongliang Luo、Md Nazmul Islam、Natalie E Sheils、John Buresh、Mackenzie Edmond- son、Peter A Merkel、Ebbing Lautenbach、Rui Duan 和 Yong Chen。面向异构临床数据的分布式学习及其在整合230个站点的COVID-19数据中的应用。《*NPJ*数字医学》，5(1):1–8，2022年。Lorenzo Trippa、Levi Waldron、Curtis Huttenhower 和 Giovanni Parmigiani。对预测方法进行贝叶斯非参数跨研究验证。《应用统计年鉴》，9(1):402–428，2015年。

Anastasios A Tsiatis。《半参数理论与缺失数据》。2006年。

Gianni Turcato、Arian Zaboli、Irena Kostic、Barbara Melchioretto、Laura Ciccariello、Eleonora Zaccaria、Alessia Olivato、Antonio Maccagnani、Norbert Pfeifer 和 Antonio Bonora。首次在急诊科评估时记录的SARS-CoV-2感染严重程度与白蛋白水平：一项多中心回顾性观察研究。《急诊医学杂志》，39(1):63–69，2022年。

Sara van deGeer、Peter B¨uhlmann、Ya'acov Ritov 和 Ruben Dezeure。关于高维模型的渐近最优置信域与检验。《统计年鉴》，42(3):1166– 1202，2014年。

Steffen Ventz、Albert Lai、Timothy F Cloughesy、Patrick Y Wen、Lorenzo Trippa 和 Brian M Alexander。利用既往临床试验和真实世界数据设计并评估外部对照组。《临床癌症研究》，25(16): 4993–5001，2019年。Thanh Vinh Vo、Trong Nghia Hoang、Young Lee 和 Tze-Yun Leong。基于观察性数据的因果效应联邦估计。*arXiv*预印本 *arXiv:2106.00456*，2021年。Cheng Wang、Larry Han、Gabriella Stein、Suzanne Day、Cedric Bien-Gund、Allison Mathews、Jason J Ong、Pei-Zhen Zhao、Shu-Fang Wei、Jennifer Walker 等。健康与医学研究中的众包：系统综述。《贫困传染病》，9(1):1–9，2020年。Ruoyu Wang、Qihua Wang 和 Wang Miao。在存在有偏来源的情况下，基于汇总统计量的稳健融合提取程序。*arXiv*预印本 *arXiv:2108.12600*，2021年。Xiaozhou Wang、Zhuoyi Yang、Xi Chen 和 Weidong Liu。线性支持向量机的分布式推断。《机器学习研究杂志》，20(113):1–41，2019年。Griffin M Weber、Chuan Hong、Zongqi Xia、Nathan P Palmer、Paul Avillach、Sehi L' Yi、Mark S Keller、Shawn N Murphy、Alba Guti´errez-Sacrist´an、Clara-Lea Bonzel 等。利用4CE协作组织的实验室指标进行国际比较，以预测COVID-19死亡率。《*NPJ*数字医学》，5(1):1–8，2022年。

Frank Windmeijer、Helmut Farbmacher、Neil Davies 和 George Davey Smith。关于使用套索回归进行含有若干无效工具变量的工具变量估计。《美国统计学会杂志》，114(527):1339–1350，2019年。Frank Windmeijer、Xiaoran Liang、Fernando P Hartwig 和 Jack Bowden。用于筛选有效工具变量的置信区间法。《英国皇家统计学会会刊：*B*辑（统计方法学）》，83(4):752–776，2021年。Zunyou Wu 和 Jennifer M McGoogan。2019年冠状病毒病（COVID-19）在中国暴发的特征与重要经验：中国疾病预防控制中心72,314例病例报告摘要。《美国医学会杂志》，323(13):1239–1242，2020年。Laure Wynants、Ben Van Calster、Gary S Collins、Richard D Riley、Georg Heinze、Ewoud Schuit、Marc MJ Bonten、Darren L Dahly、Johanna A Damen、Thomas PA Debray 等。COVID-19诊断与预后的预测模型：系统综述与批判性评价。《英国医学杂志》，369，2020年。Min-ge Xie 和 Peng Wang。适用于有限样本和大样本推断的Repro样本方法。*arXiv*预印本 *arXiv:2206.06421*，2022年。Ruoxuan Xiong、Allison Koenecke、Michael Powell、Zhu Shen、Joshua T Vogelstein 和 Susan Athey。异质性观察性数据中的联邦因果推断。*arXiv*预印本 *arXiv:2107.11732*，2021年。Cun-Hui Zhang 和 Stephanie S Zhang。高维线性模型中低维参数的置信区间。《英国皇家统计学会会刊：*B*辑（统计方法学）》，76(1):217–242，2014年。Tianqi Zhao、Guang Cheng 和 Han Liu。大规模异质数据的部分线性框架。《统计年刊》，44(4):1400，2016年。Yinchu Zhu 和 Jelena Bradic。稠密高维线性模型中的线性假设检验。《美国统计学会杂志》，113(524):1583–1600，2018年。Hui Zou。自适应套索及其预言机性质。《美国统计学会杂志》，101(476):1418–1429，2006年。

A. 补充方法与理论

A.1. RIFL 算法

算法 1 用于多源推断的抽样方法

输入： 站点特定估计量 $\{\widehat{\beta}^{(l)}, \widehat{\sigma}_l\}_{1\le l\le L}$；差异度量 $\{\widehat{\mathcal{D}}_{l,k}\}_{1\le l<k\le L}$ 及其标准误估计值 $\{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\}_{1\le l<k\le L}$；$M \ge 1$，$\rho(M) \in (0, 1)$，以及水平 ν，$\alpha > 0$。

输出： 置信区间 CI；每个站点的泛化能力度量

对于 $l \leftarrow 1$ 到 $L-1$ **执行**
  对于 $k \leftarrow l+1$ 到 L **执行**
    计算 $\widehat{\mathcal{L}}_{l,k} = \widehat{\beta}^{(l)} - \widehat{\beta}^{(k)}$ 和标准误 $\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) = \sqrt{\widehat{\sigma}_l^2 + \widehat{\sigma}_k^2}$；
  结束循环
结束循环
对于 $m \leftarrow 1$ 到 M **执行**
  对差异度量 $\{\widehat{\mathcal{D}}_l^{[m]}$ 进行重采样
  计算重采样后的检验统计量 $\{\widehat{S}_l^{[m]}$ $\widehat{\mathcal{L}}_l^{[m]}$
  按式 (14) 构造采样投票矩阵 $\widehat{H}^{[m]}$；
  按式 (15) 构造最大团 $\widehat{\mathcal{V}}^{[m]}$；
  构造聚合集 $\widehat{\mathcal{V}}^{[m]}$ 如式 (17) 所示；
  按式 (18) 构造置信区间 $\mathrm{CI}^{[m]}$；
结束循环
按照 (16) 构造索引集 $\mathcal{M}$；
返回 (19) 中定义的置信区间；
对于 $l \leftarrow 1$ 到 L **执行**
  返回可推广性度量 $\widehat{p}_l = \sum_{m\in\mathcal{M}} \mathbf{1}(l \in \widetilde{\mathcal{V}}^{[m]})/|\mathcal{M}|$；
结束循环

A.2. 高维下 $\widehat{\mathcal{D}}_{l,k}$ 的理论性质

我们在以下算法 2 中总结了所提出的高维距离度量估计量。我们将考虑高维线性模型或逻辑回归模型，并建立 $\widehat{\mathcal{D}}_{l,k}$ 的理论性质。我们回顾正文中的记号。对于 $1 \le l \le L$，定义 $\boldsymbol{\eta}^{(l)} = (\mu_l, [\theta^{(l)}]) \in \mathbb{R}^{d+1}$ 和 $\widetilde{\boldsymbol{\eta}}^{(l)} = (\widetilde{\mu}_l, [\widetilde{\theta}^{(l)}]) \in \mathbb{R}^{d+1}$ 以及 $\widetilde{X}_i^{(l)} = (1, (X_i^{(l)}))$，其中 $1 \le i \le n_l$. 令 $\widetilde{\gamma}_{l,k} := (0, [\widetilde{\theta}^{(l)} - \widetilde{\theta}^{(k)}])$，且 $\gamma_{l,k} := (0, [\theta^{(l)} - \theta^{(k)}])$，其中 $1 \le l < k \le L$。根据上下文清楚时，我们分别用 $\widetilde{\gamma}$ 和 γ 表示 $\widetilde{\gamma}_{l,k}$ 和 $\gamma_{l,k}$。

我们的理论分析基于 Guo 等人 [2021a]中的分析。我们将 Guo 等人 [2021a] 中的模型假设推广到包含 L 高维回归模型的情形。

(A1) 对于 $1 \le l \le L$，这些行 $\{\widetilde{X}_i^{(l)}\}_{1\le i\le n}$ 是独立同分布的 d维次高斯随机向量，其中 $\Sigma^{(l)} = \mathbf{E}(\widetilde{X}_i^{(l)}[\widetilde{X}_i^{(l)}])$，且 $\Sigma^{(l)}$ 满足 $c_0 \le \lambda_{\min}\left(\Sigma^{(l)}\right) \le \lambda_{\max}\left(\Sigma^{(l)}\right) \le C_0$，其中某些正数常量为 $C_0 \ge c_0 > 0$；假设高维向量 $\eta^{(l)}$ 的非零元素个数少于s。

算法 2 高维距离度量 输入：多源数据 $\{X^{(l)}, Y(l)\}_{1\le l\le L}$。输出： $\{\widehat{\beta}^{(l)}, \widehat{\text{SE}}(\widehat{\beta}^{(l)}) \widehat{\mathcal{D}}_{l,k}, \widehat{\text{SE}}(\widehat{\mathcal{D}}_{l,k})\}_{1\le l<k\le L}$1: 对 $l \leftarrow 1$ 至 L 循环 2: 按照式 (29) 计算初始的 $\widetilde{\mu}_l, \widetilde{\theta}^{(l)}$； 3: 计算满足式 (30) 的 $\widehat{\beta}^{(l)}$ 和 $\text{SE}(\widehat{\beta}^{(l)})$； 4: 循环结束 5: 将 $\{\widetilde{\mu}_l, \widetilde{\theta}^{(l)}, \widehat{\beta}^{(l)}, \widehat{\text{SE}}(\widehat{\beta}^{(l)})\}_{1\le l\le L}$ 广播至所有 L 站点； 6: 对 $l \leftarrow 1$ 至 L 循环 7: 按照式 (33) 计算投影方向 $\{\widehat{u}_k^{(l)}\}_{k\ne l}$； 8: 按照式 (34) 计算 $\{\widehat{\delta}_k^{(l)}\}_{k\ne l}$，并按照式 (35) 计算 $\{\widehat{\text{V}}_k^{(l)}\}_{k\ne l}$； 9: 循环结束 10: 按照式 (36) 计算 $\{\mathcal{D}_{l,k}\}_{1\le k<l\le L}$，并按照式 (37) 计算 $\{\text{SE}(\mathcal{D}_{l,k})\}_{1\le k<l\le L}$。

条件 (A1) 对高维协变量施加尾部条件，并假设总体二阶矩矩阵可逆。

对于高维逻辑回归，我们施加以下条件，

(A2) 以大于 $1-d^{-c}$ 的概率，$\min\{h([\widetilde{X}_i^{(l)}]^{\intercal}\ \eta^{(l)})\ 1-h([\widetilde{X}_i^{(l)}]^{\intercal}\ \eta^{(l)})\} \ge c_{\min}$ 对于 $1 \le i \le n$ 和某个较小的正数常量 $c_{\min} \in (0,\ 1)$。

施加条件 (A2)，使病例概率一致地远离 0 和 1。对于 $1 \le i \le n$ 的情形，当 $[\widetilde{X}_i^{(l)}]^{\intercal}\ \eta^{(l)}$ 有界时，条件 (A2) 以高概率成立。

我们假设惩罚最大似然估计量 $\widetilde{\theta}^{(l)}$ 满足以下性质：

(B) 以大于 $1 - d^{-c} - \exp(-cn)$ 的概率，其中 $c > 0$ 为某个常数，

$$\|\widetilde{\eta}^{(l)} - \eta^{(l)}\|_1 \le Cs\,(\log d/n)^{1/2} \quad \text{and} \quad \|\widetilde{\eta}^{(l)}_{S_l^c} - \eta^{(l)}_{S_l^c}\|_1 \le C_0\|\widetilde{\eta}^{(l)}_{S_l} - \eta^{(l)}_{S_l}\|_1$$

其中 S_l 表示 $\eta^{(l)}$ 的支撑集，$C > 0$ 和 $C_0 > 0$ 为正数常量。

对于高维线性模型，Bickel 等人的定理 7.2 [2009] 证明了 Lasso 估计量满足条件 (B)。对于高维逻辑回归，参见 Guo 等人的命题 1 [2021a]，其中给出了证明在 (29) 中定义的惩罚最大似然估计量 $\widetilde{\theta}^{(l)}$ 满足条件 (B) 的示例；另请参见其中的参考文献。

THEOREM 5. 假设高维线性回归满足条件 (A1) 和 (B)，或者高维逻辑回归满足条件 (A1)、(A2)和 (B)，则在 (33) 中定义的 $\tau_n \asymp (\log n)^{1/2}$ 满足 $\tau_n s\log d/\sqrt{n} \to 0$。对于任意常数 $0 < \alpha < 1$，在 (36) 中定义的差异度估计量 $\widehat{\mathcal{D}}_{l,k}$ 以及在 (37) 中定义的标准误估计量 $\text{SE}(\mathcal{D}_{l,k})$ 满足 (3) $1 \le l < k \le L$.

我们在 B.5 节给出上述定理的证明。

A.3. 双重稳健 *ATE* 估计量的影响函数

对于来自第 l 个源总体、样本量为 n_l 的独立同分布样本，我们用 $X_i^{(l)} \in \mathbb{R}^p$ 表示第 i 个观测的协变量向量，用 $X_{ij}^{(l)} \in \mathbb{R}$ 表示第 i 个观测中的第 j 个协变量。

我们考虑使用广义线性模型（GLM）估计结果回归函数和倾向评分的情形。具体而言，对于倾向得分模型，我们拟合以下 GLM：

$$\mathbf{E}\left[A^{(l)}|X^{(l)}\right] = h\left(\alpha_{l,0} + \sum_{j=1}^{B} \alpha_{l,j}\psi_j(X^{(l)})\right) = h\left(\boldsymbol{\alpha}_l^\top \widetilde{X}^{(l)}\right),$$

其中 $\{\psi_1(\cdot)\ldots,\psi_B(\cdot)\}$ 是一组基函数，$\widetilde{X}^{(l)} = (1, \psi_1(X^{(l)})\ldots,\psi_B(X^{(l)}))^\top$ 和 $\boldsymbol{\alpha}_l = (\alpha_{l,0}, \alpha_{l,1}, \ldots, \alpha_{l,B})$。一个简单的例子是令 $\psi_j(X_i^{(l)}) = X_{ij}^{(l)}$ 对于 $j \in \{1,\ldots,p\}$，这样我们便得到一个包含每个协变量主效应的 GLM。我们通过求解以下估计方程来估计系数：

$$\frac{1}{n_l}\sum_{i=1}^{n_l} \widetilde{X}_i^{(l)}\left\{A_i^{(l)} - h\left(\boldsymbol{\alpha}^\top \widetilde{X}_i^{(l)}\right)\right\} = 0,$$

并得到一个估计系数，记为 $\widehat{\boldsymbol{\alpha}}_l$。

对于结局回归，我们在每个来源站点的每个治疗组内拟合广义线性模型：

$$\mathbf{E}(Y^{(l)} \mid A^{(l)} = a, X^{(l)}) = g\left(\boldsymbol{\gamma}_{a,0}^{(l)} + \sum_{j=1}^{B_\gamma} \boldsymbol{\gamma}_{a,j}^{(l)}\phi_j(X^{(l)})\right) = g\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l)}\right), \quad \text{for } a \in \{0,1\},$$

其中 $W^{(l)} = (1, \phi_1(X^{(l)}) \ldots\ldots, \phi_{B_\gamma}(X^{(l)}))$ 表示一组基函数 $\{\phi_1(\cdot) \ldots\ldots, \phi_{B_\gamma}(\cdot)\}$，且 $\boldsymbol{\gamma}_a^{(l)} = (\boldsymbol{\gamma}_{a\,0}^{(l)}, \boldsymbol{\gamma}_{a\,1}^{(l)}, \ldots\ldots, \boldsymbol{\gamma}_{a\,B_\gamma}^{(l)})$。令 $I\{\cdot\}$ 表示指示函数。我们通过求解以下估计方程来估计系数 $\boldsymbol{\gamma}_a^{(l)}$：

$$\frac{1}{n_l}\sum_{i=1}^{n_l} I\left\{A_i^{(l)} = a\right\} W_i^{(l)}\left\{Y_i^{(l)} - g\left([W_i^{(l)}]^\top \boldsymbol{\gamma}_a^{(l)}\right)\right\} = 0,$$

我们将该估计值记为 $\widehat{\boldsymbol{\gamma}}_a^{(l)}$。

对于密度比，我们考虑指数倾斜模型 $\omega_l(X^{(l)}; \boldsymbol{\eta}^{(l)}) = \exp([\boldsymbol{\eta}^{(l)}]^\top \widetilde{W}^{(l)})$，其中 $\widetilde{W}^{(l)} = (1, \varphi_1(X^{(l)}) \ldots\ldots, \varphi_{B_\omega}(X^{(l)}))^\top$ 表示一组基函数 $\{\varphi_1(\cdot) \ldots\ldots, \varphi_{B_\omega}(\cdot)\}$。我们通过求解以下估计方程来估计参数 $\boldsymbol{\eta}^{(l)}$

$$\frac{1}{n_l}\sum_{i=1}^{n_l} \exp\left([\boldsymbol{\eta}^{(l)}]^\top \widetilde{W}_i^{(l)}\right)\widetilde{W}_i^{(l)} = \frac{1}{N}\sum_{j=1}^{N} \widetilde{W}_j^{\mathcal{T}}/N$$

其中 $\widetilde{W}_j^{\mathcal{T}} = (1, \varphi_1(X_j^{\mathcal{T}})\ldots, \varphi_{B_\omega}(X_j^{\mathcal{T}}))^\top$，$X_j^{\mathcal{T}}$ 表示目标数据集中的第 j 个观测值，其中 $j \in \{1,\ldots,N\}$。我们将得到的估计记为 $\widehat{\boldsymbol{\eta}}^{(l)}$。本质上，我们估计

参数 $\boldsymbol{\eta}^{(l)}$，方法是匹配源数据集和规模大得多的目标数据集中一组基函数的样本均值。

令 $\boldsymbol{\alpha}_l^*$、$\boldsymbol{\gamma}_{,0}^{(l)\ *}$、$\boldsymbol{\gamma}_{,1}^{(l)\ *}$ 和 $\boldsymbol{\eta}^{(l),*}$ 分别表示 $\widehat{\boldsymbol{\alpha}}_l$、$\widehat{\boldsymbol{\gamma}}_0^{(l)}$、$\widehat{\boldsymbol{\gamma}}_1^{(l)}$ 和 $\widehat{\boldsymbol{\eta}}^{(l)}$ 的概率极限。这些总体参数的精确定义见定理6的证明。定义以下矩阵:

$$
\begin{aligned}
C_{\boldsymbol{\alpha},l} &= \mathbf{E}_l\left[\widetilde{X}^{(l)} h'\left((\boldsymbol{\alpha}_l^*)^\top \widetilde{X}^{(l)}\right)\left(\widetilde{X}^{(l)}\right)^\top\right];\\
C_{\boldsymbol{\gamma}_a,l} &= \mathbf{E}_l\left[I\left\{A^{(l)}=a\right\} W^{(l)} g'\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l),*}\right)[W^{(l)}]^\top\right], \quad a\in\{0,1\};\\
C_{\boldsymbol{\eta},l} &= -\mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right)\widetilde{W}^{(l)}[\widetilde{W}^{(l)}]^\top\right],
\end{aligned}
$$

其中，$\mathbf{E}_l$ 表示相对于 l 个源总体中 $(X^{(l)}, A^{(l)}, Y_{(l)})$ 的联合分布的期望。

回顾第 5.3 节，目标平均处理效应估计量的形式为 $\widehat{\theta}^{(l)} = \widehat{M}^{(l)} + \widehat{\delta}^{(l)}$ 其中

$$
\widehat{M}^{(l)} = \frac{1}{N}\sum_{i=1}^{N}\left\{m(1, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_1^{(l)}) - m(0, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_0^{(l)})\right\}
$$

and

$$
\widehat{\delta}^{(l)} = \frac{1}{n_l}\sum_{i=1}^{n_l}\omega_l(X_i^{(l)}; \widehat{\boldsymbol{\eta}}^{(l)})\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1} I\{A_i^{(l)}=a\}}{\pi_l(a, X_i^{(l)}; \widehat{\boldsymbol{\alpha}}_l)}\{Y_i^{(l)} - m(A_i^{(l)}, X_i^{(l)}; \widehat{\boldsymbol{\gamma}}_a^{(l)})\}\right\}.
$$

在这种情况下，我们有

$$
\begin{aligned}
m(a, X_i^{(l)}; \widehat{\boldsymbol{\gamma}}_a^{(l)}) &= g\left([W_i^{(l)}]^\top \widehat{\boldsymbol{\gamma}}_a^{(l)}\right);\\
m(a, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_a^{(l)}) &= g\left([W_i^{\mathcal{T}}]^\top \widehat{\boldsymbol{\gamma}}_a^{(l)}\right);\\
\omega_l(X_i^{(l)}; \widehat{\boldsymbol{\eta}}^{(l)}) &= \exp\left([\widehat{\boldsymbol{\eta}}^{(l)}]^\top \widetilde{W}_i^{(l)}\right);\\
\pi_l(a, X_i^{(l)}; \widehat{\boldsymbol{\alpha}}_l) &= h\left((\widehat{\boldsymbol{\alpha}}_l)^\top \widetilde{X}_i^{(l)}\right)^a\left\{1 - h\left((\widehat{\boldsymbol{\alpha}}_l)^\top \widetilde{X}_i^{(l)}\right)\right\}^{(1-a)}.
\end{aligned}
$$

为便于记号，我们定义函数 $\tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(\cdot)$ 和 $\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(\cdot)$，使得

$$
\tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(x^{(l)}) = m(1, x^{(l)}; \boldsymbol{\gamma}_1) - m(0, x^{(l)}; \boldsymbol{\gamma}_0),
$$

and

$$
\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(x^{(l)}, a^{(l)}, y^{(l)}) = \omega_l(x^{(l)}; \boldsymbol{\eta})\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1} I\{a^{(l)}=a\}}{\pi_l(a, x^{(l)}; \boldsymbol{\alpha})}\{y^{(l)} - m(a, x^{(l)}; \boldsymbol{\gamma}_a\}\right\}.
$$

这些函数将在以下推导以及 $\widehat{\theta}^{(l)}$ 的影响函数中频繁出现。还请注意，函数 τ 和 ξ 由干扰模型参数索引。

定义以下量:

$$\boldsymbol{d}_{\boldsymbol{\gamma}_0} = -\mathbf{E}_{\mathcal{T}}\left[g'\left([W^{\mathcal{T}}]^\top \boldsymbol{\gamma}_0^{(l),*}\right) W^{\mathcal{T}}\right] + \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right) \frac{I\{A^{(l)}=0\}}{1-h\left([\widetilde{X}^{(l)}]^\top \boldsymbol{\alpha}_l^*\right)} \left\{g'\left([W^{(l)}]^\top \boldsymbol{\gamma}_0^{(l),*}\right) W^{(l)}\right\}\right];$$

$$\boldsymbol{d}_{\boldsymbol{\gamma}_1} = \mathbf{E}_{\mathcal{T}}\left[g'\left([W^{\mathcal{T}}]^\top \boldsymbol{\gamma}_1^{(l),*}\right) W^{\mathcal{T}}\right] - \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right) \frac{I\{A^{(l)}=1\}}{h\left([\widetilde{X}^{(l)}]^\top \boldsymbol{\alpha}_l^*\right)} \left\{g'\left([W^{(l)}]^\top \boldsymbol{\gamma}_1^{(l),*}\right) W^{(l)}\right\}\right];$$

$$\boldsymbol{d}_{\boldsymbol{\eta}} = \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right) \widetilde{W}^{(l)} \left\{\sum_{a=0}^{1} \frac{(-1)^{a+1} I\{A^{(l)}=a\}}{\pi_l(a, X^{(l)}; \boldsymbol{\alpha}_l^*)} \left\{Y^{(l)} - g\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l),*}\right)\right\}\right\}\right];$$

$$\boldsymbol{d}_{\boldsymbol{\alpha}} = \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right) \left\{\sum_{a=0}^{1} \frac{-I\{A^{(l)}=a\}\pi_l'(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}{\pi_l^2(a, X^{(l)}; \boldsymbol{\alpha}_l^*)} \left\{Y^{(l)} - g\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l),*}\right)\right\}\right\} \widetilde{X}^{(l)}\right].$$

THEOREM 6. 当 $N \gg n_l$时，目标平均处理效应估计量 $\widehat{\theta}^{(l)}$ 渐近线性，其影响函数 τ_θ 满足

$$\begin{aligned}
\tau_\theta(x^{(l)}, a^{(l)}, y^{(l)}) = {} & \xi_{\boldsymbol{\eta}^{(l),*}, \boldsymbol{\alpha}_l^*, \boldsymbol{\gamma}_0^{(l),*}, \boldsymbol{\gamma}_1^{(l),*}}(x^{(l)}, a^{(l)}, y^{(l)}) \\
& + \boldsymbol{d}_{\boldsymbol{\gamma}_0}^\top C_{\boldsymbol{\gamma}_0, l}^{-1} I\left\{a^{(l)}=0\right\} w^{(l)} \left\{y^{(l)} - g\left([w^{(l)}]^\top \boldsymbol{\gamma}_0^{(l),*}\right)\right\} \\
& + \boldsymbol{d}_{\boldsymbol{\gamma}_1}^\top C_{\boldsymbol{\gamma}_1, l}^{-1} I\left\{a^{(l)}=1\right\} w^{(l)} \left\{y^{(l)} - g\left([w^{(l)}]^\top \boldsymbol{\gamma}_1^{(l),*}\right)\right\} \\
& + \boldsymbol{d}_{\boldsymbol{\eta}}^\top C_{\boldsymbol{\eta}, l}^{-1} \left\{\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{w}^{(l)}\right) \widetilde{w}^{(l)} - \mathbf{E}_{\mathcal{T}}\left[\widetilde{W}^{\mathcal{T}}\right]\right\} \\
& + \boldsymbol{d}_{\boldsymbol{\alpha}}^\top C_{\boldsymbol{\alpha}, l}^{-1} \widetilde{x}^{(l)} \left\{a^{(l)} - h\left((\boldsymbol{\alpha}_l^*)^\top \widetilde{x}^{(l)}\right)\right\}.
\end{aligned}$$

这里，向量 $\widetilde{x}^{(l)}$、$w^{(l)}$ 和 $\widetilde{w}^{(l)}$ 表示由协变量向量 $x^{(l)}$ 导出的基函数向量。

上述定理的证明见 B.6节。

$\widehat{\theta}^{(l)}$ 的方差可通过在源数据上对函数 $\widehat{\tau}_\theta$ 计算经验方差来相合估计，其中 $\widehat{\tau}_\theta$ 是通过将干扰模型参数（$\boldsymbol{\eta}^{(l),*}$，$\boldsymbol{\alpha}_l^*$，$\boldsymbol{\gamma}_0^{(l),*}$，$\boldsymbol{\gamma}_1^{(l),*}$）、偏导数（$\boldsymbol{d}_{\boldsymbol{\gamma}_0}$，$\boldsymbol{d}_{\boldsymbol{\gamma}_1}$，$\boldsymbol{d}_{\boldsymbol{\eta}}$，$\boldsymbol{d}_{\boldsymbol{\alpha}}$）以及矩阵 $C_{\boldsymbol{\gamma}_0, l}$、$C_{\boldsymbol{\gamma}_1, l}$、$C_{\boldsymbol{\eta}, l}$ 和 $C_{\boldsymbol{\alpha}, l}$ 的估计值代入而得到的 τ_θ 的估计值。这些偏导数和矩阵的相合估计量在定理6的证明中给出。

B. 证明

B.1. 定理*1*的证明

用 $\mathcal{O}$ 表示观测数据。我们为数据 $\mathcal{O}$ 定义以下事件，

$$\mathcal{E} = \left\{\max_{1\le l<k\le L} \max\left\{\frac{\left|\widehat{\mathcal{D}}_{l,k} - \mathcal{D}_{l,k}\right|}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}, \frac{\left|\widehat{\mathcal{L}}_{l,k} - \mathcal{L}_{l,k}\right|}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right\} \le z_{\nu/[2L(L-1)]}\right\}. \tag{41}$$

由于站点特定估计量 $\left\{\widehat{\beta}^{(l)},\ \widehat{\sigma}_l\right\}_{1\le l\le L}$ 满足 (2)，我们得出

$$\limsup_{n\to\infty}\mathbf{P}\left(\left|\widehat{\mathcal{L}}_{l,k}-\mathcal{L}_{l,k}\right|/\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})\ge z_\alpha\right)\le\alpha\quad\text{for}\quad 0<\alpha<1. \tag{42}$$

由于差异性度量 $\{\mathcal{D}_{l,k},\mathrm{SE}(\mathcal{D}_{l,k})\}_{1\le l<k\le L}$ 满足式 (3)，我们应用并集界，得到

$$\liminf_{n\to\infty}\mathbf{P}(\mathcal{E})\ge 1-\nu. \tag{43}$$

接下来，我们针对给定的样本量 n进行理论分析，然后考察关于 n.的极限。我们将堆叠向量 $\widehat{U}$, $\{U^{[m]}\}_{1\le m\le M}\in\mathbb{R}^{L(L-1)}$定义为

$$\widehat{U}=\begin{pmatrix}\left\{\frac{\widehat{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right\}_{1\le l<k\le L}\\ \left\{\frac{\widehat{\mathcal{L}}_{l,k}-\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right\}_{1\le l<k\le L}\end{pmatrix}\quad\text{and}\quad U^{[m]}=\begin{pmatrix}\left\{\frac{\widehat{\mathcal{D}}_{l,k}-\mathcal{D}^{[m]}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right\}_{1\le l<k\le L}\\ \left\{\frac{\widehat{\mathcal{L}}_{l,k}-\mathcal{L}^{[m]}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right\}_{1\le l<k\le L}\end{pmatrix}.$$

回顾一下， U是观测数据 $\mathcal{O}$.的函数。令 $f(\cdot\mid\mathcal{O})$ 表示给定数据 $\mathcal{O}$时 $U^{[m]}$的条件密度函数，即

$$f(U^{[m]}=U\mid\mathcal{O})=\prod_{1\le j\le L(L-1)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{U_j^2}{2}\right).$$

在事件 $\mathcal{E}$上，我们有

$$\sum_{1\le j\le L(L-1)}\frac{\widehat{U}_j^2}{2}\le\frac{L(L-1)}{2}\left[z_{\nu/[2L(L-1)]}\right]^2,$$

并进一步得到

$$f(U^{[m]}=\widehat{U}\mid\mathcal{O})\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\ge c(\nu):=\left(\frac{1}{\sqrt{2\pi}}\right)^{L(L-1)}\exp\left(-\frac{L(L-1)}{2}\left[z_{\nu/[2L(L-1)]}\right]^2\right). \tag{44}$$

我们用 $\mathbf{P}(\cdot\mid\mathcal{O})$ 表示相对于观测数据 $\mathcal{O}$ 的条件概率。注意，

$$\begin{aligned}&\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\\&=1-\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\ge\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\\&=1-\prod_{m=1}^{M}\left[1-\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\right]\\&\ge 1-\exp\left[-\sum_{m=1}^{M}\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\right],\end{aligned}$$

其中，第二个等式由给定数据 $\mathcal{O}$ 时 $\{U^{[m]}\}_{1\leq m\leq M}$ 的条件独立性得出，最后一个不等式则由 $1-x\leq e^{-x}$. 得出。应用上述不等式，我们得到

$$\begin{aligned}
&\mathbf{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
&\geq\left(1-\exp\left[-\sum_{m=1}^{M}\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\right]\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
&=1-\exp\left[-\sum_{m=1}^{M}\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\right].
\end{aligned}\tag{45}$$

在证明的其余部分，我们为.....建立一个下界

$$\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}},\tag{46}$$

并应用（45）为.....建立一个下界

$$\mathbf{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right).$$

我们进一步将（46）中的目标概率分解为

$$\begin{aligned}
&\mathbf{P}\left(\|U^{[m]}-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
=&\int f(U^{[m]}=U\mid\mathcal{O})\cdot\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
=&\int f(U^{[m]}=\widehat{U}\mid\mathcal{O})\cdot\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
&+\int[f(U^{[m]}=U\mid\mathcal{O})-f(U^{[m]}=\widehat{U}\mid\mathcal{O})]\cdot\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}.
\end{aligned}\tag{47}$$

根据（44），我们得到

$$\begin{aligned}
&\int f(U^{[m]}=\widehat{U}\mid\mathcal{O})\cdot\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
&\geq c(\nu)\cdot\int\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\leq \mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\
&\geq c(\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}.
\end{aligned}\tag{48}$$

存在 $t\in(0,\ 1)$ 使得

$$f(U^{[m]}=U\mid\mathcal{O})-f(U^{[m]}=\widehat{U}\mid\mathcal{O})=[\triangledown f(\widehat{U}+t(U-\widehat{U}))]^\intercal(U-\widehat{U}),$$

with

$$\triangledown f(u)=\left(\frac{1}{\sqrt{2\pi}}\left[-u\cdot\exp\left(-\frac{u^2}{2}\right)\right]\right)^{L(L-1)}.$$

由于 $\|\nabla f\|_2$ 有上界，存在一个正数常量 $C>0$ 使得

$$\left|f(U^{[m]}=U\mid\mathcal{O})-f(U^{[m]}=\widehat{U}\mid\mathcal{O})\right|\le C\sqrt{L(L-1)}\|U-\widehat{U}\|_\infty.$$

接下来我们证明

$$\begin{aligned}&\left|\int[f(U^{[m]}=U\mid\mathcal{O})-f(U^{[m]}=\widehat{U}\mid\mathcal{O})]\cdot\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\right|\\&\le C\sqrt{L(L-1)}\cdot\mathrm{err}_n(M,\nu)\cdot\int\mathbf{1}_{\left\{\|U-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\right\}}dU\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\&=C\sqrt{L(L-1)}\cdot\mathrm{err}_n(M,\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}.\end{aligned}\tag{49}$$

由于 $\mathrm{err}_n(M,\ \nu)\ \to 0$ 和 $c(\nu)$ 是一个正数常量，因此存在一个正整数 M_0，使得

$$C\sqrt{L(L-1)}\cdot\mathrm{err}_n(M,\nu)\le\frac{1}{2}c(\nu)\quad\text{for}\quad M\ge M_0.$$

W我们结合上述不等式、(47)、(48) 和 (49)，得到对于 $M\ge M_0$,

$$\mathbf{P}\left(\|U-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\ge\frac{1}{2}c(\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}.$$

结合 (45)，我们证明对于 $M\ge M_0$,

$$\begin{aligned}&\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\\&\ge1-\exp\left[-M\cdot\frac{1}{2}c(\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\right]\\&=\left(1-\exp\left[-M\cdot\frac{1}{2}c(\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\right]\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}.\end{aligned}\tag{50}$$

令 $\mathbf{E}_{\mathcal{O}}$ 表示关于观测数据 $\mathcal{O}$，的期望，我们进一步关于 $\mathcal{O}$ 积分，并证明对于 $M\ge M_0$,

$$\begin{aligned}&\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\right)\\&=\mathbf{E}_{\mathcal{O}}\left[\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\right]\\&\ge\mathbf{E}_{\mathcal{O}}\left[\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\mid\mathcal{O}\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\right]\\&\ge\mathbf{E}_{\mathcal{O}}\left[\left(1-\exp\left[-M\cdot\frac{1}{2}c(\nu)\cdot[2\mathrm{err}_n(M,\nu)]^{L(L-1)}\right]\right)\cdot\mathbf{1}_{\mathcal{O}\in\mathcal{E}}\right].\end{aligned}$$

根据定义 $\mathrm{err}_n(M,\nu)\ =\frac{1}{2}\left[{}^{2}\log c\ \nu\quad{}^{n}\ ()M\right]^{L}\ \_\_\_(L-1),$，我们证明对于 $M\ge M_0$,

$$\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le\mathrm{err}_n(M,\nu)\right)\ge(1-n^{-1})\cdot\mathbf{P}\left(\mathcal{E}\right).$$

我们进一步应用（43）并证明

$$\liminf_{n\to\infty}\lim_{M\to\infty}\mathbf{P}\left(\min_{1\le m\le M}\|U^{[m]}-\widehat{U}\|_\infty\le \mathrm{err}_n(M,\nu)\right)\ge\mathbf{P}\left(\mathcal{E}\right)\ge 1-\nu.$$

B.2. 定理 *2* 的证明
在接下来的内容中，我们首先通过固定样本量 n 来进行分析。定义分离 n

$$\mathcal{L}_{\min}(n)=\min_{l\in\mathcal{V}^c}|\beta^{(l)}-\beta^*|.$$

注意， $\mathcal{L}_{\min}(n)$ 是样本量 n 的函数，并且可能随着样本量 n. 的增大而趋于零。我们定义事件

$$\mathcal{E}_1=\left\{\min_{1\le m\le M}\max_{1\le l<k\le L}\max\left\{\left|\frac{\widehat{\mathcal{D}}_{l,k}^{[m]}-\mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right|,\left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m]}-\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\right\}\le \mathrm{err}_n(M,\nu)\right\}.$$

在事件 $\mathcal{E}_1$ 上，我们用 m^* 表示满足下列条件的索引

$$\max_{1\le l<k\le L}\max\left\{\left|\frac{\widehat{\mathcal{D}}_{l,k}^{[m^*]}-\mathcal{D}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}\right|,\left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]}-\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\right\}\le \mathrm{err}_n(M,\nu). \tag{51}$$

回顾一下， $T=z_{\nu/[2L(L-1)]}$. 注意到

$$\mathbf{P}\left(\theta^*\in\mathrm{CI}\right)\ge\mathbf{P}\left(\{\theta^*\in\mathrm{CI}\}\cap\mathcal{E}_1\right)\ge\mathbf{P}\left(\left\{\theta^*\in\mathrm{CI}^{[m^*]}\right\}\cap\mathcal{E}_1\right). \tag{52}$$

我们首先研究 $\widehat{\mathcal{V}}^{[m^*]}$ 和 $\widetilde{\mathcal{V}}^{[m^*]}$，的性质，这些性质对后续证明很有用。如果 $\rho(M)\cdot T\ge$ err$_n(M,$ $\nu)$，那么 $\mathcal{V}(\theta^*)$ 在图 $\mathcal{G}([L],\ \widehat{H}^{[m^*]})$ 中构成一个团，而最大团 $\widehat{\mathcal{V}}^{[m^*]}$ 满足

$$\left|\widehat{\mathcal{V}}^{[m^*]}\right|\ge|\mathcal{V}(\theta^*)|>L/2.$$

上述不等式意味着 $m^*\in\mathcal{M}$，并且存在 $1\le l\le L$ 使得

$$l\in\mathcal{V}\cap\widehat{\mathcal{V}}^{[m^*]}. \tag{53}$$

我们引入以下引理来量化式 (17) 中定义的集合 $\widetilde{\mathcal{V}}^{[m^*]}$ 。以下引理的证明推迟到 B.2.1 节。

Lemma 1. 假设事件 $\mathcal{E}_1$ 成立，且索引 m^* 满足 (51)。若索引 l,k 满足 $\widehat{H}_l^{[m^*]}{}_{,\ k}=1$，则有

$$\left|\frac{\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|\le\rho(M)\cdot T+\mathrm{err}_n(M,\nu). \tag{54}$$

此外，如果 $\rho(M)\cdot T\ge$ 误差$_n(M,\nu)$ 且

$$2\rho(M)\cdot T\cdot\max_{j_1\in\mathcal{V},j_2\in\mathcal{V}^c}\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{j_1,j_2})<\mathcal{L}_{\min}(n) \tag{55}$$

then

$$\widetilde{\mathcal{V}}^{[m^*]} = \mathcal{V}(\theta^*). \tag{56}$$

现在我们分析 $\mathbf{P}\left(\left\{\theta^* \in \text{置信区间}^{[m^*]}\right\} \cap \mathcal{E}_1\right)$，然后通过应用 (52) 建立覆盖性质。对于给定的 n，由于 $\rho(M) \to 0$，存在 M_n，使得若 $M \geq M_n$, $\rho(M) \cdot T \geq$ 误差$_n(M,\ \nu)$ 且 (55) 成立。因此，对于 $M \geq M_n$，我们有

$$\begin{aligned}\mathbf{P}\left(\left\{\theta^* \in \mathrm{CI}^{[m^*]}\right\} \cap \mathcal{E}_1\right) &= \mathbf{P}\left(\left\{\left|\frac{\sum_{l\in\mathcal{V}}\left(\widehat{\beta}^{(l)} - \beta^*\right)/\widehat{\sigma}_l^2}{\sqrt{\sum_{l\in\mathcal{V}} 1/\widehat{\sigma}_l^2}}\right| \leq z_{\alpha_1/2}\right\} \cap \mathcal{E}_1\right) \\ &\geq \mathbf{P}\left(\left|\frac{\sum_{l\in\mathcal{V}}\left(\widehat{\beta}^{(l)} - \beta^*\right)/\widehat{\sigma}_l^2}{\sqrt{\sum_{l\in\mathcal{V}} 1/\widehat{\sigma}_l^2}}\right| \leq z_{\alpha_1/2}\right) - \left[1 - \mathbf{P}\left(\mathcal{E}_1\right)\right].\end{aligned}$$

对 M 取极限，可得

$$\liminf_{M\to\infty} \mathbf{P}\left(\left\{\theta^* \in \mathrm{CI}^{[m^*]}\right\} \cap \mathcal{E}_1\right) \geq \mathbf{P}\left(\left|\frac{\sum_{l\in\mathcal{V}}\left(\widehat{\beta}^{(l)} - \beta^*\right)/\widehat{\sigma}_l^2}{\sqrt{\sum_{l\in\mathcal{V}} 1/\widehat{\sigma}_l^2}}\right| \leq z_{\alpha_1/2}\right) - 1 + \liminf_{M\to\infty} \mathbf{P}\left(\mathcal{E}_1\right).$$

结合定理 1 以及满足 (2) 的 $\left\{\widehat{\beta}^{(l)},\ \widehat{\sigma}_l\right\}_{1\leq l\leq L}$，我们得到定理 2。

B.2.1. 引理 *1* 的证明**(54) 的证明**。对于 $\widehat{H}_l^{[m^*]}{}_{,\ k} = 1$,
我们应用 (14)，得到

$$\left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right| \leq \rho(M) \cdot T. \tag{57}$$

我们应用 (51)，得到

$$\left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]} - \mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right| \leq \mathrm{err}_n(M, \nu). \tag{58}$$

我们通过应用 (57) 和 (58)，并利用三角不等式，得到 (54)

$$\left|\frac{\mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right| \leq \left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]} - \mathcal{L}_{l,k}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right| + \left|\frac{\widehat{\mathcal{L}}_{l,k}^{[m^*]}}{\widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k})}\right|.$$

（56）的证明。对于 $k \in \mathcal{V}$，我们结合（51）与条件 $\rho(M) \cdot T \geq \mathrm{err}_n(M,\ \nu)$，从而得到 $\widehat{H}_k^{[m^*]}{}_{,\ j} = 1$，对任意 $j \in \mathcal{V}$. 。根据多数规则，我们有 $\left\|\widehat{H}_k^{[m^*]}{}_{,\ \cdot}\right\|_0 > L/2$，对于 $k \in \mathcal{V}$.

对于 $k \notin \mathcal{V}$，我们将（54）与条件（55）结合使用，并建立 $\widehat{H}_k^{[m^*]}{}_{,\ j} = 0$ 对于 $j \in \mathcal{V}$.根据多数规则，我们有 $\left\|\widehat{H}_k^{[m^*]}{}_{,\ \cdot}\right\|_0 < L/2$ 对于 $k \notin \mathcal{V}$. 因此，$\widetilde{\mathcal{V}}^{[m^*]} = \mathcal{V}$.

B.3. 定理*3*的证明
回顾 $\mathcal{V}^* = \mathcal{V}(\theta^*)$。我们定义事件

$$\mathcal{E}_2 = \left\{ \left| \widehat{\mathcal{L}}_{l,k}^{[m]} - \widehat{\mathcal{L}}_{l,k} \right| / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \leq \sqrt{2\log n + 2\log M} \right\}$$
$$\mathcal{E}_3 = \left\{ \left| \mathcal{L}_{l,k} - \widehat{\mathcal{L}}_{l,k} \right| / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \leq \sqrt{2\log n} \right\}.$$

由 (11) 和 (42)，应用并集界可得

$$\lim_{n\to\infty} \lim_{M\to\infty} \mathbf{P}(\mathcal{E}_2 \cap \mathcal{E}_3) = 1. \tag{59}$$

根据三角不等式，有

$$\left| \widehat{\mathcal{L}}_{l,k}^{[m]} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) - \mathcal{L}_{l,k} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \right| \leq \left| \widehat{\mathcal{L}}_{l,k}^{[m]} - \widehat{\mathcal{L}}_{l,k} \right| / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) + \left| \mathcal{L}_{l,k} - \widehat{\mathcal{L}}_{l,k} \right| / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \tag{60}$$

在事件 $\mathcal{E}_2 \cap \mathcal{E}_3$, 上，有

$$\left| \widehat{\mathcal{L}}_{l,k}^{[m]} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) - \mathcal{L}_{l,k} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \right| \leq 2\sqrt{2\log n + 2\log M}. \tag{61}$$

良分离条件 (25) 意味着，对于 $l \in \mathcal{V}$ 和 $k \in \mathcal{V}^c$,

$$\left| \mathcal{L}_{l,k} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \right| > 2\sqrt{2\log n + 2\log M} + \rho(M) \cdot T.$$

在事件 $\mathcal{E}_2 \cap \mathcal{E}_3$，若 $l \in \mathcal{V}$ 和 $k \in \mathcal{V}^c$ 满足（61），则

$$\left| \widehat{\mathcal{L}}_{l,k}^{[m]} / \widehat{\mathrm{SE}}(\widehat{\mathcal{L}}_{l,k}) \right| > \rho(M) \cdot T.$$

这等价于 $\widehat{H}_l^{[m]}{}_{,\ k} = 0$. 也就是说， $k \notin \mathcal{V}^{[m]}$ 对于 $m \in \mathcal{M}$.

上述推导表明，若所有索引 $k \in \mathcal{V}^c$ 均满足良好分离条件（61），则对于 $m \in \mathcal{M}$. 有 $\mathcal{V}^{[m]} \subset \mathcal{V}^*$。由于 $|\mathcal{V}^{[m]}| > L/2$ 和 $|\mathcal{V}^*| = \lfloor L/2 \rfloor + 1$，对于$m \in \mathcal{M}$. 有 $\mathcal{V}^{[m]} = \mathcal{V}^*$。这意味着 $\mathbf{P}$（置信区间 $=$ 置信区间$_{\mathrm{ora}}$）$\geq \mathbf{P}$（$\mathcal{E}_2 \cap \mathcal{E}_3$）。由此，定理由（59）得证。

B.4. 定理*4*的证明
定义

$$\mathrm{SE}(\widehat{\mathcal{D}}_{l,k}) = \sqrt{4\gamma^\intercal C^{(l)}\gamma/n_l + 4\gamma^\intercal C^{(k)}\gamma/n_k + 1/\min\{n_l, n_k\}}$$

and

$$\mathrm{SE}_0(\widehat{\mathcal{D}}_{l,k}) = \sqrt{4\gamma^\intercal C^{(l)}\gamma/n_l + 4\gamma^\intercal C^{(k)}\gamma/n_k}.$$

由于 $\widehat{\theta}^{(l)}$ 和 $\widehat{C}^{(l)}$ 分别是 $\theta^{(l)}$ 和 $C^{(l)}$的相合估计量，我们有

$$\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k}) / \mathrm{SE}(\widehat{\mathcal{D}}_{l,k}) \xrightarrow{d} 1. \tag{62}$$

定义 $c_n =$(1/最小值$\{n_l, n_k\}$)$^{1/4}$. 由（26）可得

$$\limsup_{n\to\infty} \mathbf{P}\left(\|\widehat{\gamma} - \gamma\|_2^2 / \widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k}) \geq c_n z_\alpha \right) = 0. \tag{63}$$

根据分解（27），我们有

$$\begin{aligned}
&\mathbf{P}\left(\left|\widehat{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}\right|/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\geq z_\alpha\right)\\
&\leq \mathbf{P}\left(|2\langle\widehat{\gamma}-\gamma,\gamma\rangle|/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\geq(1-c_n)z_\alpha\right)+\mathbf{P}\left(\|\widehat{\gamma}-\gamma\|_2^2/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\geq c_n z_\alpha\right)\\
&\leq \mathbf{P}\left(|2\langle\widehat{\gamma}-\gamma,\gamma\rangle|/\mathrm{SE}_0(\widehat{\mathcal{D}}_{l,k})\geq(1-c_n)z_\alpha\cdot\frac{\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})}{\mathrm{SE}(\widehat{\mathcal{D}}_{l,k})}\right)+\mathbf{P}\left(\|\widehat{\gamma}-\gamma\|_2^2/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\geq c_n z_\alpha\right),
\end{aligned}\tag{64}$$

其中，第一个不等式由并集界得出，第二个不等式则由 $\widehat{\ }$ $\widehat{\ }$关系 $\mathrm{SE}(\mathcal{D}_{l,k})\geq\mathrm{SE}_0(\mathcal{D}_{l,k})$。结合上述分解、(62)、(63) 以及

$$\frac{\langle\widehat{\gamma}-\gamma,\gamma\rangle}{\sqrt{4\gamma^\intercal C^{(l)}\gamma/n_l+4\gamma^\intercal C^{(k)}\gamma/n_l}}\xrightarrow{d}N(0,1).$$

B.5. 定理 *5* 的证明

以下分解对于构造误差分量的相合估计量至关重要：对于任意向量 $u\in\mathbb{R}^d$,

$$\begin{aligned}
&[\widehat{u}_k^{(l)}]^\intercal\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}W_i^{(l)}\widetilde{X}_i^{(l)}(Y_i-h([\widetilde{X}_i^{(l)}]^\intercal\widetilde{\eta}^{(l)}))-\langle\eta^{(l)}-\widetilde{\eta}^{(l)},\widetilde{\gamma}\rangle\\
&=\left(\widehat{\Sigma}^{(l)}\widehat{u}_k^{(l)}-\widetilde{\gamma}\right)^\intercal(\eta^{(l)}-\widetilde{\eta}^{(l)})+[\widehat{u}_k^{(l)}]^\intercal\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}W_i^{(l)}\epsilon_i^{(l)}\widetilde{X}_i^{(l)}+[\widehat{u}_k^{(l)}]^\intercal\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}\Delta_i^{(l)}\widetilde{X}_i^{(l)},
\end{aligned}\tag{65}$$

其中 $\widehat{\Sigma}^{(l)}=\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}\widetilde{X}_i^{(l)}\left[\widetilde{X}_i^{(l)}\right]^\intercal$，近似误差 $\Delta_i^{(l)}$ 定义为

$$\Delta_i^{(l)}=W_i^{(l)}\cdot\int_0^1(1-t)h''([\widetilde{X}_i^{(l)}]^\intercal\widetilde{\eta}^{(l)}+t[\widetilde{X}_i^{(l)}]^\intercal[\eta^{(l)}-\widetilde{\eta}^{(l)}])dt\cdot([\widetilde{X}_i^{(l)}]^\intercal[\eta^{(l)}-\widetilde{\eta}^{(l)}])^2.\tag{66}$$

注意到 f 或对于带有 $h(x)=x$ 的线性结果模型，近似误差 $\Delta_i^{(l)}=0$. Defin e

$$\widetilde{\mathcal{D}}_{l,k}=\|\widetilde{\theta}^{(l)}-\widetilde{\theta}^{(k)}\|_2^2+2\widehat{\delta}_k^{(l)}-2\widehat{\delta}_l^{(k)}.$$

由于 $\mathcal{D}_{l,\ k}\geq 0$，我们有

$$\left|\widehat{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}\right|\leq\left|\widetilde{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}\right|.$$

要证明 (3)，只需证明

$$\limsup_{n\to\infty}\mathbf{P}\left(\left|\widetilde{\mathcal{D}}_{l,k}-\mathcal{D}_{l,k}\right|/\widehat{\mathrm{SE}}(\widehat{\mathcal{D}}_{l,k})\geq z_\alpha\right)\leq\alpha\quad\text{for}\quad 0<\alpha<1\tag{67}$$

其中，z_α 表示标准正态分布的上分位数。

下面我们证明（67）。应用（31）和（65），得到

$$\begin{aligned}
&\widetilde{\mathcal{D}}_{l,k} - \mathcal{D}_{l,k} = -\|\widetilde{\gamma} - \gamma\|_2^2 \\
&+ \left(\widehat{\Sigma}^{(l)}\widehat{u}_k^{(l)} - \widetilde{\gamma}\right)^{\intercal}(\eta^{(l)} - \widetilde{\eta}^{(l)}) + [\widehat{u}_k^{(l)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}} W_i^{(l)}\epsilon_i^{(l)}\widetilde{X}_i^{(l)} + [\widehat{u}_k^{(l)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}\Delta_i^{(l)}\widetilde{X}_i^{(l)} \\
&+ \left(\widehat{\Sigma}^{(k)}\widehat{u}_l^{(k)} - \widetilde{\gamma}\right)^{\intercal}(\eta^{(k)} - \widetilde{\eta}^{(k)}) + [\widehat{u}_l^{(k)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(k)}|}\sum_{i\in\mathcal{S}_2^{(k)}} W_i^{(k)}\epsilon_i^{(k)}\widetilde{X}_i^{(k)} + [\widehat{u}_l^{(k)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(k)}|}\sum_{i\in\mathcal{S}_2^{(k)}}\Delta_i^{(k)}\widetilde{X}_i^{(k)}.
\end{aligned} \tag{68}$$

我们应用 Guo 等人中的引理 5 [2021a] 并得到

$$\frac{[\widehat{u}_k^{(l)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}} W_i^{(l)}\epsilon_i^{(l)}\widetilde{X}_i^{(l)} + [\widehat{u}_l^{(k)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(k)}|}\sum_{i\in\mathcal{S}_2^{(k)}} W_i^{(k)}\epsilon_i^{(k)}\widetilde{X}_i^{(k)}}{\sqrt{\widehat{\mathrm{V}}_k^{(l)} + \widehat{\mathrm{V}}_l^{(k)}}} \xrightarrow{d} N(0,1) \tag{69}$$

with

$$\widehat{\mathrm{V}}_k^{(l)} = \left(\widehat{u}_k^{(l)}\right)^{\intercal}\left[\frac{1}{|\mathcal{S}_2^{(l)}|^2}\sum_{i\in\mathcal{S}_2^{(l)}} W_i^{(l)}\widetilde{X}_i^{(l)}[\widetilde{X}_i^{(l)}]^{\intercal}\right]\widehat{u}_k^{(l)},$$

and

$$\widehat{\mathrm{V}}_l^{(k)} = \left(\widehat{u}_l^{(k)}\right)^{\intercal}\left[\frac{1}{|\mathcal{S}_2^{(k)}|^2}\sum_{i\in\mathcal{S}_2^{(k)}} W_i^{(k)}\widetilde{X}_i^{(k)}[\widetilde{X}_i^{(k)}]^{\intercal}\right]\widehat{u}_l^{(k)}.$$

根据条件 (B)，我们有

$$\|\widetilde{\gamma} - \gamma\|_2^2 \le C\frac{s\log d}{\min\{n_l, n_k\}}. \tag{70}$$

我们应用 Guo 等人中的 (22) [2021a] 并得到

$$\left|\left(\widehat{\Sigma}^{(l)}\widehat{u}_k^{(l)} - \widetilde{\gamma}\right)^{\intercal}(\eta^{(l)} - \widetilde{\eta}^{(l)})\right| \le C\frac{s\log d}{n_l}, \quad \left|\left(\widehat{\Sigma}^{(k)}\widehat{u}_l^{(k)} - \widetilde{\gamma}\right)^{\intercal}(\eta^{(k)} - \widetilde{\eta}^{(k)})\right| \le C\frac{s\log d}{n_k}. \tag{71}$$

我们应用 Guo 等人中的 (23) [2021a] 并得到

$$\begin{aligned}
\left|[\widehat{u}_k^{(l)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(l)}|}\sum_{i\in\mathcal{S}_2^{(l)}}\Delta_i^{(l)}\widetilde{X}_i^{(l)}\right| &\le C\tau_n\frac{s\log d}{n_l} \\
\left|[\widehat{u}_l^{(k)}]^{\intercal}\frac{1}{|\mathcal{S}_2^{(k)}|}\sum_{i\in\mathcal{S}_2^{(k)}}\Delta_i^{(k)}\widetilde{X}_i^{(k)}\right| &\le C\tau_n\frac{s\log d}{n_k}
\end{aligned} \tag{72}$$

我们通过结合分解（68）、渐近极限（69）、误差界（70）、（71）、（72）以及条件 $\tau_n s\log d/\sqrt{n} \to 0$.，建立（67）。

B.6. 定理*6*的证明

我们令 $\boldsymbol{\alpha}_l^*$ 表示 $\widehat{\boldsymbol{\alpha}}_l$ 的概率极限，它满足以下方程:

$$\mathbf{E}_l\left[\widetilde{X}^{(l)}\left\{A^{(l)}-h\left((\boldsymbol{\alpha}_l^*)^\top\widetilde{X}^{(l)}\right)\right\}\right]=0,$$

其中，$\mathbf{E}_l$ 表示相对于第l个源总体中$(X^{(l)},A^{(l)},Y_{(l)})$联合分布的期望。根据估计方程理论， $\widehat{\boldsymbol{\alpha}}_l$ 具有如下渐近线性展开:

$$\widehat{\boldsymbol{\alpha}}_l-\boldsymbol{\alpha}_l^*=\frac{1}{n_l}\sum_{i=1}^{n_l}C_{\boldsymbol{\alpha},l}^{-1}\widetilde{X}_i^{(l)}\left\{A_i^{(l)}-h\left((\boldsymbol{\alpha}_l^*)^\top\widetilde{X}_i^{(l)}\right)\right\}+o_P(n_l^{-1/2}),$$

其中，矩阵$C_{\boldsymbol{\alpha},l}$ 定义为

$$C_{\boldsymbol{\alpha},l}=\mathbf{E}_l\left[\widetilde{X}^{(l)}h'\left((\boldsymbol{\alpha}_l^*)^\top\widetilde{X}^{(l)}\right)\left(\widetilde{X}^{(l)}\right)^\top\right].$$

注意，$C_{\boldsymbol{\alpha},l}$ 可以通过以下方式相合估计:

$$\widehat{C}_{\boldsymbol{\alpha},l}=\frac{1}{n_l}\sum_{i=1}^{n_l}\widetilde{X}_i^{(l)}h'\left((\widehat{\boldsymbol{\alpha}}_l)^\top\widetilde{X}_i^{(l)}\right)\left(\widetilde{X}_i^{(l)}\right)^\top.$$

令 $\boldsymbol{\gamma}^{(l)}{}_{,a}{}^*$表示 $\widehat{\boldsymbol{\gamma}}_a^{(l)}$的概率极限，它满足以下方程:

$$\mathbf{E}_l\left[I\left\{A^{(l)}=a\right\}W^{(l)}\left\{Y^{(l)}-g\left([W^{(l)}]^\top\boldsymbol{\gamma}_a^{(l),*}\right)\right\}\right]=0.$$

估计量 $\widehat{\boldsymbol{\gamma}}_a^{(l)}$具有如下渐近线性展开:

$$\widehat{\boldsymbol{\gamma}}_a^{(l)}-\boldsymbol{\gamma}_a^{(l),*}=\frac{1}{n_l}\sum_{i=1}^{n_l}C_{\boldsymbol{\gamma}_a,l}^{-1}I\left\{A_i^{(l)}=a\right\}W_i^{(l)}\left\{Y_i^{(l)}-g\left([W_i^{(l)}]^\top\boldsymbol{\gamma}_a^{(l),*}\right)\right\}+o_P(n_l^{-1/2}),$$

其中，矩阵 $C_{\boldsymbol{\gamma}_a,\ l}$ 定义为

$$C_{\boldsymbol{\gamma}_a,l}=\mathbf{E}_l\left[I\left\{A^{(l)}=a\right\}W^{(l)}g'\left([W^{(l)}]^\top\boldsymbol{\gamma}_a^{(l),*}\right)[W^{(l)}]^\top\right].$$

矩阵 $C_{\boldsymbol{\gamma}_a,\ l}$ 可通过其经验版本进行相合估计，

$$\widehat{C}_{\boldsymbol{\gamma}_a,l}=\frac{1}{n_l}\sum_{i=1}^{n_l}I\left\{A_i^{(l)}=a\right\}W_i^{(l)}g'\left([W_i^{(l)}]^\top\widehat{\boldsymbol{\gamma}}_a^{(l)}\right)[W_i^{(l)}]^\top.$$

令 $\boldsymbol{\eta}^{(l),*}$ 表示 $\widehat{\boldsymbol{\eta}}^{(l)}$ 的概率极限，使得

$$\mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top\widetilde{W}^{(l)}\right)\widetilde{W}^{(l)}\right]=\mathbf{E}_{\mathcal{T}}\left[\widetilde{W}^{\mathcal{T}}\right],$$

其中 $\mathbf{E}_{\mathcal{T}}$ 表示相对于目标总体中协变量分布的期望算子。根据估计方程的标准理论，估计量 $\widehat{\boldsymbol{\eta}}^{(l)}$ 再次具有渐近线性，其展开式如下:

$$\widehat{\boldsymbol{\eta}}^{(l)} - \boldsymbol{\eta}^{(l),*} = C_{\boldsymbol{\eta},l}^{-1} \left\{ \frac{1}{n_l} \sum_{i=1}^{n_l} \exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}_i^{(l)} \right) \widetilde{W}_i^{(l)} - \frac{1}{N} \sum_{j=1}^{N} \widetilde{W}_j^{\mathcal{T}} \right\} + o_P(n_l^{-1/2}),$$

其中，矩阵$C_{\boldsymbol{\eta},\ l}$ 定义为

$$C_{\boldsymbol{\eta},l} = -\mathbf{E}_l \left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)} \right) \widetilde{W}^{(l)} [\widetilde{W}^{(l)}]^\top \right],$$

且可以通过以下方式相合估计

$$\widehat{C}_{\boldsymbol{\eta},l} = -\frac{1}{n_l} \sum_{i=1}^{n_l} \exp\left([\widehat{\boldsymbol{\eta}}^{(l)}]^\top \widetilde{W}_i^{(l)} \right) \widetilde{W}_i^{(l)} [\widetilde{W}_i^{(l)}]^\top.$$

注意，当 $N \gg n_l$ 时， $\widehat{\boldsymbol{\eta}}^{(l)}$ 的渐近线性展开简化为

$$\widehat{\boldsymbol{\eta}}^{(l)} - \boldsymbol{\eta}^{(l),*} = \frac{1}{n_l} \sum_{i=1}^{n_l} C_{\boldsymbol{\eta},l}^{-1} \left\{ \exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}_i^{(l)} \right) \widetilde{W}_i^{(l)} - \mathbf{E}_{\mathcal{T}} \left[\widetilde{W}^{\mathcal{T}} \right] \right\} + o_P(n_l^{-1/2}).$$

现在，我们推导站点特异性目标平均处理效应估计量的渐近线性展开。回顾第 5.3 节，目标平均处理效应估计量的形式为 $\widehat{\theta}^{(l)} = \widehat{M}^{(l)} + \widehat{\delta}^{(l)}$ 其中

$$\widehat{M}^{(l)} = \frac{1}{N} \sum_{i=1}^{N} \left\{ m(1, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_1^{(l)}) - m(0, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_0^{(l)}) \right\}$$

and

$$\widehat{\delta}^{(l)} = \frac{1}{n_l} \sum_{i=1}^{n_l} \omega_l(X_i^{(l)}; \widehat{\boldsymbol{\eta}}^{(l)}) \left\{ \sum_{a=0}^{1} \frac{(-1)^{a+1} I\{A_i^{(l)} = a\}}{\pi_l(a, X_i^{(l)}; \widehat{\boldsymbol{\alpha}}_l)} \{Y_i^{(l)} - m(A_i^{(l)}, X_i^{(l)}; \widehat{\boldsymbol{\gamma}}_a^{(l)})\} \right\}.$$

在这种情况下，我们有

$$\begin{aligned}
m(a, X_i^{(l)}; \widehat{\boldsymbol{\gamma}}_a^{(l)}) &= g\left([W_i^{(l)}]^\top \widehat{\boldsymbol{\gamma}}_a^{(l)} \right); \\
m(a, X_i^{\mathcal{T}}; \widehat{\boldsymbol{\gamma}}_a^{(l)}) &= g\left([W_i^{\mathcal{T}}]^\top \widehat{\boldsymbol{\gamma}}_a^{(l)} \right); \\
\omega_l(X_i^{(l)}; \widehat{\boldsymbol{\eta}}^{(l)}) &= \exp\left([\widehat{\boldsymbol{\eta}}^{(l)}]^\top \widetilde{W}_i^{(l)} \right); \\
\pi_l(a, X_i^{(l)}; \widehat{\boldsymbol{\alpha}}_l) &= h\left((\widehat{\boldsymbol{\alpha}}_l)^\top \widetilde{X}_i^{(l)} \right)^a \left\{ 1 - h\left((\widehat{\boldsymbol{\alpha}}_l)^\top \widetilde{X}_i^{(l)} \right) \right\}^{(1-a)}.
\end{aligned}$$

另外请回顾，我们定义以下函数$\tau_{\boldsymbol{\gamma}_0,\ \boldsymbol{\gamma}_1}(\cdot)$和$\xi_{\boldsymbol{\eta},\ \boldsymbol{\alpha},\ \boldsymbol{\gamma}_0,\ \boldsymbol{\gamma}_1}(\cdot)$，使得

$$\tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(x^{(l)}) = m(1, x^{(l)}; \boldsymbol{\gamma}_1) - m(0, x^{(l)}; \boldsymbol{\gamma}_0),$$

and

$$\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}(x^{(l)},a^{(l)},y^{(l)})=\omega_l(x^{(l)};\boldsymbol{\eta})\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1}I\{a^{(l)}=a\}}{\pi_l(a,x^{(l)};\boldsymbol{\alpha})}\{y^{(l)}-m(a,x^{(l)};\boldsymbol{\gamma}_a\}\right\}.$$

我们介绍处理经验过程时经常使用的以下记号：对于一般函数 f_1 和 f_2，定义 $P_l f_1=\int f_1 d\mathrm{P}_l$，其中$\mathrm{P}_l$ 表示第 l个源总体中（$X^{(l)}$、 $A^{(l)}$、 $Y(l)$）的联合分布；定义 $P_{\mathcal{T}} f_2=\int f_2 d\mathrm{P}_{\mathcal{T}}$，其中$\mathrm{P}_{\mathcal{T}}$表示目标总体中的协变量分布。此外，令$P_{n,l}f_1=\sum_{i=1}^{n_l} f_1(X_i^{(l)},A_i^{(l)},Y(l)_i)/n_l$ 表示 f_1 在第 l个源数据上的经验平均值，令$P_{n,\mathcal{T}}f_2=\sum_{i=1}^{N} f_2(X_i^{\mathcal{T}})/N$ 表示 f_2 在目标数据上的经验平均值。借助这些记号，我们现在可以对 $\widehat{\theta}^{(l)}$进行线性化。

首先，我们注意到估计量 $\widehat{\theta}^{(l)}$ 可以写成

$$\widehat{\theta}^{(l)}=P_{n,\mathcal{T}}\left(\tau_{\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}\right)+P_{n,l}\left(\xi_{\widehat{\boldsymbol{\eta}}^{(l)},\widehat{\boldsymbol{\alpha}}_l,\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}\right).$$

因此，估计量 $\widehat{\theta}^{(l)}$ 具有如下展开式

$$\begin{aligned}
\widehat{\theta}^{(l)}-\theta^{(l)}&=P_{n,\mathcal{T}}\left(\tau_{\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}\right)-P_{\mathcal{T}}\left(\tau_{\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)\\
&\quad+P_{n,l}\left(\xi_{\widehat{\boldsymbol{\eta}}^{(l)},\widehat{\boldsymbol{\alpha}}_l,\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}\right)-P_l\left(\xi_{\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)\\
&=(P_{n,\mathcal{T}}-P_{\mathcal{T}})\left(\tau_{\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)+P_{\mathcal{T}}\left(\tau_{\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}-\tau_{\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)\\
&\quad+(P_{n,l}-P_l)\left(\xi_{\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)+P_l\left(\xi_{\widehat{\boldsymbol{\eta}}^{(l)},\widehat{\boldsymbol{\alpha}}_l,\widehat{\boldsymbol{\gamma}}_0^{(l)},\widehat{\boldsymbol{\gamma}}_1^{(l)}}-\xi_{\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)+o_P(n_l^{-1/2})\\
&=(P_{n,\mathcal{T}}-P_{\mathcal{T}})\left(\tau_{\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)+(P_{n,l}-P_l)\left(\xi_{\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}\right)+o_P(n_l^{-1/2})\\
&\quad+\left\{P_{\mathcal{T}}\left(\frac{\partial\tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\gamma}_0}|_{(\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)+P_l\left(\frac{\partial\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\gamma}_0}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)\right\}^{\top}\left(\widehat{\boldsymbol{\gamma}}_0^{(l)}-\boldsymbol{\gamma}_0^{(l),*}\right)\\
&\quad+\left\{P_{\mathcal{T}}\left(\frac{\partial\tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\gamma}_1}|_{(\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)+P_l\left(\frac{\partial\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\gamma}_1}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)\right\}^{\top}\left(\widehat{\boldsymbol{\gamma}}_1^{(l)}-\boldsymbol{\gamma}_1^{(l),*}\right)\\
&\quad+\left\{P_l\left(\frac{\partial\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\eta}}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)\right\}^{\top}\left(\widehat{\boldsymbol{\eta}}^{(l)}-\boldsymbol{\eta}^{(l),*}\right)\\
&\quad+\left\{P_l\left(\frac{\partial\xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial\boldsymbol{\alpha}}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right)\right\}^{\top}(\widehat{\boldsymbol{\alpha}}_l-\boldsymbol{\alpha}_l^*).
\end{aligned}$$

当目标数据中的样本量 N满足 $N\gg n_l$时，项 $(P_{n,\mathcal{T}}-P_{\mathcal{T}})\tau_{\boldsymbol{\gamma}^{(l),0*},\boldsymbol{\gamma}^{(l),1*}}$的阶为 $o_P(n_l^{-1/2})$，因此可以忽略。偏导数可以显式计算。

具体而言，

$$\begin{aligned}
\boldsymbol{d}_{\boldsymbol{\gamma}_0} &= P_{\mathcal{T}}\left(\frac{\partial \tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\gamma}_0}|_{(\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) + P_l\left(\frac{\partial \xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\gamma}_0}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) \\
&= -\mathbf{E}_{\mathcal{T}}\left[\frac{\partial m}{\partial \boldsymbol{\gamma}_0}(0, X^{\mathcal{T}}; \boldsymbol{\gamma}_0^{(l),*})\right] + \mathbf{E}_l\left[\omega_l(X^{(l)}; \boldsymbol{\eta}^{(l),*})\frac{I\{A^{(l)}=0\}}{\pi_l(0, X^{(l)}; \boldsymbol{\alpha}_l^*)}\left\{\frac{\partial m}{\partial \boldsymbol{\gamma}_0}(0, X^{(l)}; \boldsymbol{\gamma}_0^{(l),*})\right\}\right] \\
&= -\mathbf{E}_{\mathcal{T}}\left[g'\left([W^{\mathcal{T}}]^\top \boldsymbol{\gamma}_0^{(l),*}\right) W^{\mathcal{T}}\right] + \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right)\frac{I\{A^{(l)}=0\}}{1-h\left([\widetilde{X}^{(l)}]^\top \boldsymbol{\alpha}_l^*\right)}\left\{g'\left([W^{(l)}]^\top \boldsymbol{\gamma}_0^{(l),*}\right) W^{(l)}\right\}\right].
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{d}_{\boldsymbol{\gamma}_1} &= P_{\mathcal{T}}\left(\frac{\partial \tau_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\gamma}_1}|_{(\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) + P_l\left(\frac{\partial \xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\gamma}_1}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) \\
&= \mathbf{E}_{\mathcal{T}}\left[\frac{\partial m}{\partial \boldsymbol{\gamma}_1}(1, X^{\mathcal{T}}; \boldsymbol{\gamma}_1^{(l),*})\right] - \mathbf{E}_l\left[\omega_l(X^{(l)}; \boldsymbol{\eta}^{(l),*})\frac{I\{A^{(l)}=1\}}{\pi_l(1, X^{(l)}; \boldsymbol{\alpha}_l^*)}\left\{\frac{\partial m}{\partial \boldsymbol{\gamma}_1}(1, X^{(l)}; \boldsymbol{\gamma}_1^{(l),*})\right\}\right] \\
&= \mathbf{E}_{\mathcal{T}}\left[g'\left([W^{\mathcal{T}}]^\top \boldsymbol{\gamma}_1^{(l),*}\right) W^{\mathcal{T}}\right] - \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right)\frac{I\{A^{(l)}=1\}}{h\left([\widetilde{X}^{(l)}]^\top \boldsymbol{\alpha}_l^*\right)}\left\{g'\left([W^{(l)}]^\top \boldsymbol{\gamma}_1^{(l),*}\right) W^{(l)}\right\}\right].
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{d}_{\boldsymbol{\eta}} &= P_l\left(\frac{\partial \xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\eta}}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) \\
&= \mathbf{E}_l\left[\frac{\partial \omega_l}{\partial \boldsymbol{\eta}}(X^{(l)}; \boldsymbol{\eta}^{(l),*})\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1} I\{A^{(l)}=a\}}{\pi_l(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}\{Y^{(l)} - m(a, X^{(l)}; \boldsymbol{\gamma}_a^{(l),*})\}\right\}\right] \\
&= \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right)\widetilde{W}^{(l)}\left\{\sum_{a=0}^{1}\frac{(-1)^{a+1} I\{A^{(l)}=a\}}{\pi_l(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}\left\{Y^{(l)} - g\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l),*}\right)\right\}\right\}\right].
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{d}_{\boldsymbol{\alpha}} &= P_l\left(\frac{\partial \xi_{\boldsymbol{\eta},\boldsymbol{\alpha},\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_1}}{\partial \boldsymbol{\alpha}}|_{(\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*})}\right) \\
&= \mathbf{E}_l\left[\omega_l(X^{(l)}; \boldsymbol{\eta}^{(l),*})\left\{\sum_{a=0}^{1}\frac{-I\{A^{(l)}=a\}\pi'(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}{\pi^2(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}\left\{Y^{(l)} - m(a, X^{(l)}; \boldsymbol{\gamma}_a^{(l),*})\right\}\right\}\right] \\
&= \mathbf{E}_l\left[\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{W}^{(l)}\right)\left\{\sum_{a=0}^{1}\frac{-I\{A^{(l)}=a\}\pi_l'(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}{\pi_l^2(a, X^{(l)}; \boldsymbol{\alpha}_l^*)}\left\{Y^{(l)} - g\left([W^{(l)}]^\top \boldsymbol{\gamma}_a^{(l),*}\right)\right\}\right\}\widetilde{X}^{(l)}\right].
\end{aligned}$$

这些偏导数可以通过用相应的样本均值替代期望，并用相合估计量 $\widehat{\boldsymbol{\eta}}^{(l)}, \widehat{\boldsymbol{\alpha}}_l, \widehat{\boldsymbol{\gamma}}_0^{(l)}$ 和 $\widehat{\boldsymbol{\gamma}}_1^{(l)}$ 分别替代未知总体参数 $\boldsymbol{\eta}^{(l),*}, \boldsymbol{\alpha}_l^*, \boldsymbol{\gamma}^{(l)}{}_{,0}{}^{*}$ 和 $\boldsymbol{\gamma}^{(l)}{}_{,1}{}^{*}$ 来估计。

由于 $\widehat{\boldsymbol{\eta}}^{(l)}, \widehat{\boldsymbol{\alpha}}_l \widehat{\boldsymbol{\gamma}}_0^{(l)}$ 和 $\widehat{\boldsymbol{\gamma}}_1^{(l)}$ 都是渐近线性的，我们为 $\widehat{\theta}^{(l)}$ 推导的展开式意味着

$\widehat{\theta}^{(l)}$ 也具有渐近线性，其影响函数为 τ_θ ，满足

$$\begin{aligned}\tau_\theta(x^{(l)},a^{(l)},y^{(l)}) = \xi_{\boldsymbol{\eta}^{(l),*},\boldsymbol{\alpha}_l^*,\boldsymbol{\gamma}_0^{(l),*},\boldsymbol{\gamma}_1^{(l),*}}(x^{(l)},a^{(l)},y^{(l)}) \\ + \boldsymbol{d}_{\boldsymbol{\gamma}_0}^\top C_{\boldsymbol{\gamma}_0,l}^{-1} I\left\{a^{(l)}=0\right\} w^{(l)}\left\{y^{(l)} - g\left([w^{(l)}]^\top \boldsymbol{\gamma}_0^{(l),*}\right)\right\} \\ + \boldsymbol{d}_{\boldsymbol{\gamma}_1}^\top C_{\boldsymbol{\gamma}_1,l}^{-1} I\left\{a^{(l)}=1\right\} w^{(l)}\left\{y^{(l)} - g\left([w^{(l)}]^\top \boldsymbol{\gamma}_1^{(l),*}\right)\right\} \\ + \boldsymbol{d}_{\boldsymbol{\eta}}^\top C_{\boldsymbol{\eta},l}^{-1}\left\{\exp\left([\boldsymbol{\eta}^{(l),*}]^\top \widetilde{w}^{(l)}\right)\widetilde{w}^{(l)} - \mathbf{E}_{\mathcal{T}}\left[\widetilde{W}^{\mathcal{T}}\right]\right\} \\ + \boldsymbol{d}_{\boldsymbol{\alpha}}^\top C_{\boldsymbol{\alpha},l}^{-1}\widetilde{x}^{(l)}\left\{a^{(l)} - h\left((\boldsymbol{\alpha}_l^*)^\top \widetilde{x}^{(l)}\right)\right\}.\end{aligned}$$

这里，向量 $\widetilde{x}^{(l)}$、 $w^{(l)}$ 和 $\widetilde{w}^{(l)}$ 表示由协变量向量 $x^{(l)}$导出的基函数向量。

C. 附加数值结果

C.1. 8 个多数站点及采用80% 规则的 *RIFL* 的模拟结果

现在，我们展示有 8 个多数站点时的模拟结果，并在 RIFL 中采用 80%规则，该规则利用了关于多数站点数量的先验信息。具体而言，我们修改式 (16) 中索引集的定义，并定义

$$\mathcal{M}_{80\%} \coloneqq \left\{1 \le m \le M : |\widehat{\mathcal{V}}^{[m]}| \ge 0.8L\right\}. \tag{73}$$

采用 80%规则的 RIFL 的定义与原始 RIFL 相同，只是将 $\mathcal{M}$替换为 $\mathcal{M}_{80\%}$。

首先，我们展示第5.1节低维预测示例中的结果。对于 $l \in\{1,2,\ldots,8\}$，我们设定 $\theta^{(l)} = \theta^* = (0.5, 0.5, 0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0, 0)$。对于 $\theta^{(9)}$和 $\theta^{(10)}$，其最后5个系数与 θ^* 相同，但前5个系数分别改为 $0.5-0.3a$ 和$0.5-0.1a$，其中 $a \in \{1,2,3,4,5\}$ 控制多数站点与非多数站点之间的分离程度。模拟设置的其他所有方面均与第6.1节相同。

我们在图C1中展示了各种方法所得95%置信区间的经验覆盖率和平均长度。我们观察到，RIFL和采用80%规则的RIFL在所有设置下都达到名义覆盖率。此外，当分离程度较高时，采用80%规则的RIFL所得区间明显短于原始RIFL置信区间，且与OBA区间长度相当。所有其他方法的表现与第6.1节中的情况相似。

接下来，我们展示第5.2节中高维预测示例的更多模拟结果，针对 θ_{11}^* 使用8个多数站点。模拟设置与第6.2节相同，但 $\theta^{(l)} = \theta^*$用于 $l \in \{1,\ldots,8\}$，而 $\theta_j^{(9)} = \theta_j^* + 0.2 + 0.05a$和 $\theta_j^{(10)} = \theta_j^* + 0.15 + 0.05a$用于 $6 \le j \le 11$，其中 a 在 $\{1,2,3,4,5\}$ 中变化，以表示多数站点与非多数站点之间不同程度的分离。结果见图C2，我们观察到与使用6个多数站点时相似的模式。θ_8^* 的模拟结果分别见图C3和图C4，对应6个和8个多数站点。同样，我们观察到与 θ_{11}^*情形相似的模式。

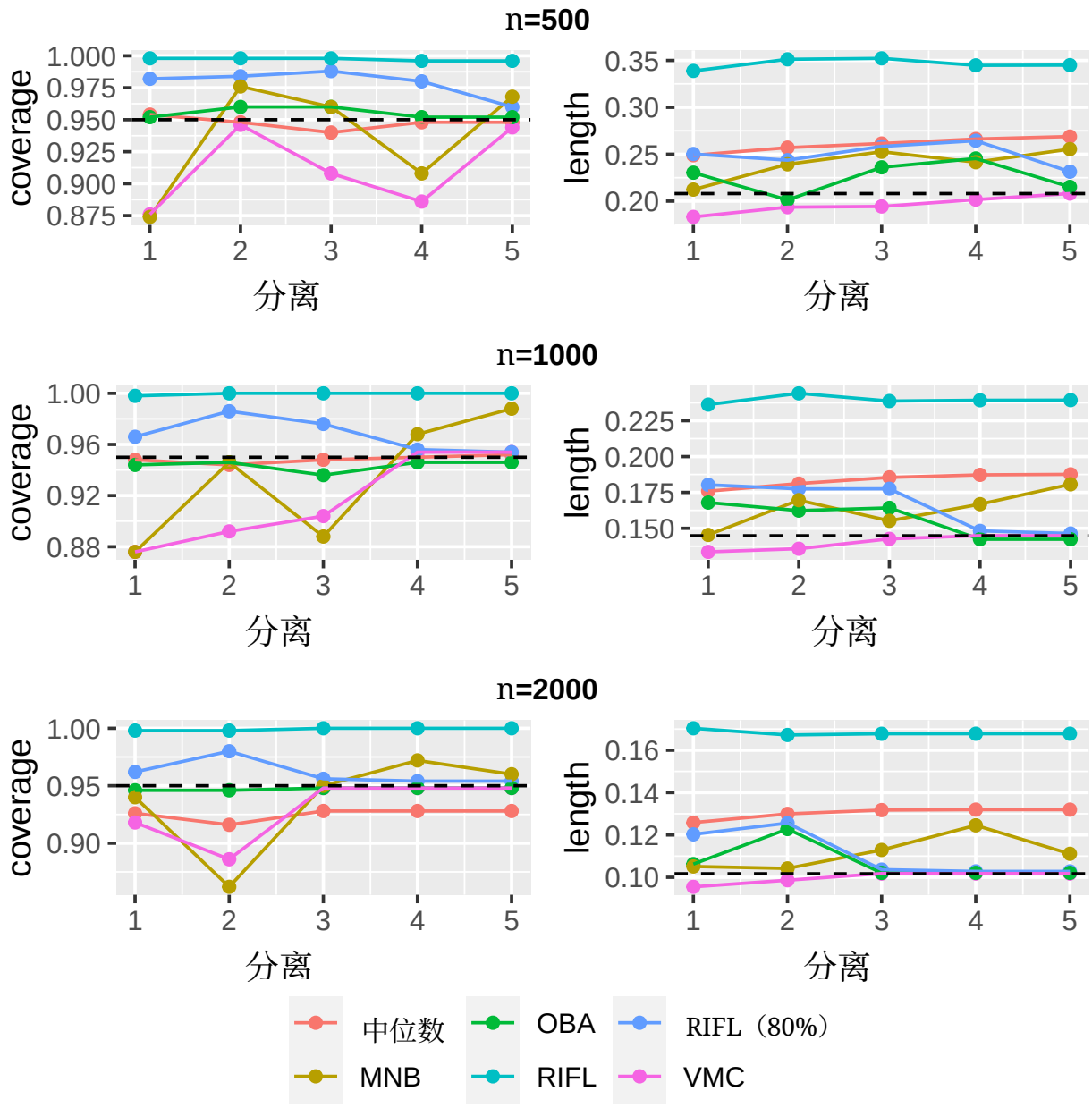

图 C1：低维预测：在 8 个多数站点下，$\beta^* = \theta_1^*$ 的 95% 置信区间覆盖率和长度随分离程度变化的情况，其中 1 表示最低（最难检测），5 表示最高（最易检测）。“median”表示基于中位数估计量的置信区间，“MNB”表示式 (38) 中的 m-out-of-n 自助法置信区间，“VMC”表示式 (10) 中的最大团投票估计量及其相应的置信区间，“OBA”表示式 (39) 中的预言机偏差感知置信区间，“RIFL”表示式 (19) 中我们提出的置信区间，“RIFL (80%)”表示利用 80%规则的 RIFL 置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

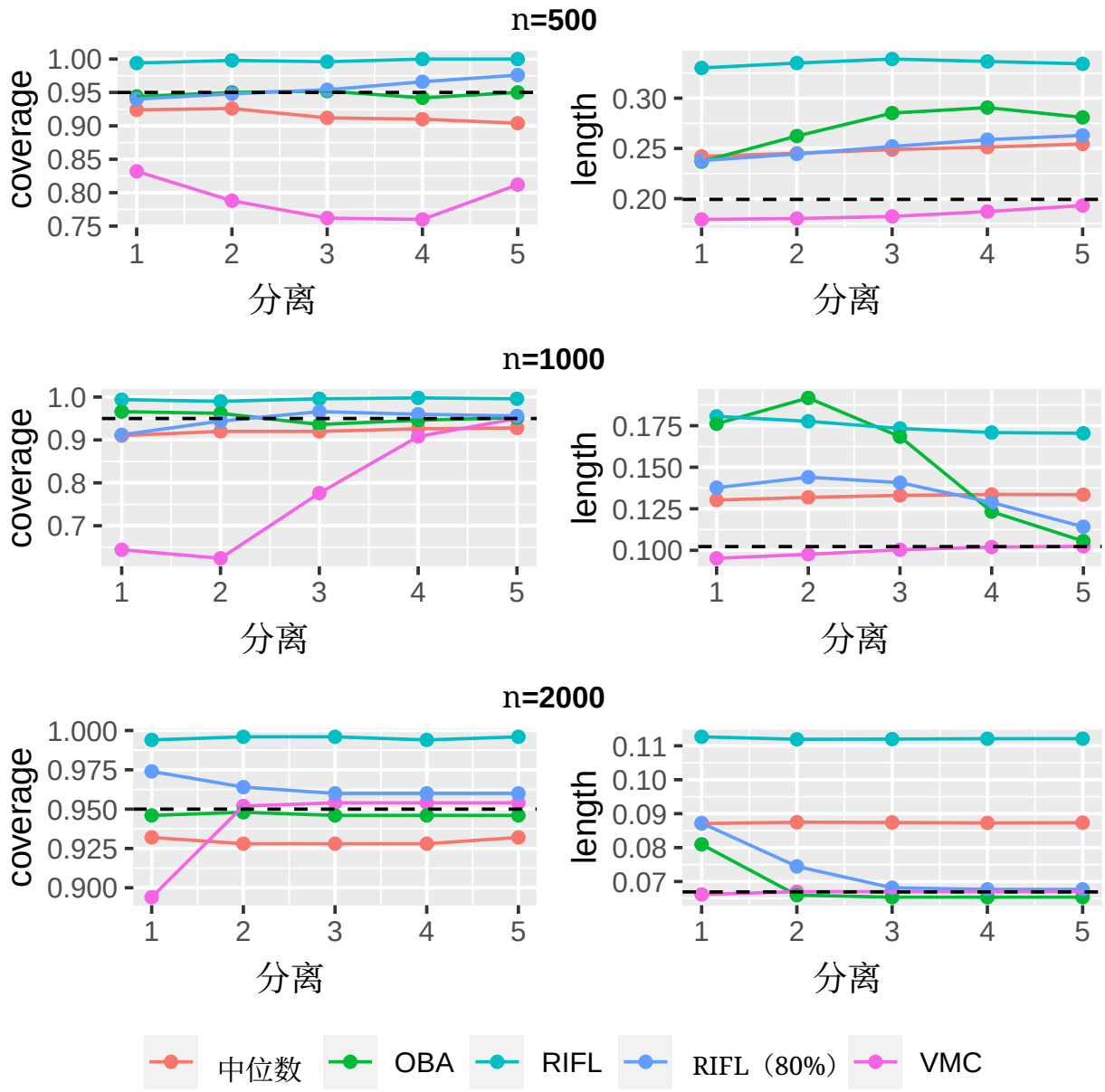

图 C2：高维预测：$\beta^* = \theta^*_{11}$ 在 8 个多数站点及不同分离程度下 95%置信区间的覆盖率和长度，其中 1 表示最低（最难检测），5 表示最高（最易检测）。“median”表示基于中位数估计量的置信区间，“VMC”表示式（10）中的最大团投票估计量及其相应的置信区间，“OBA”表示式（39）中的预言机偏差感知置信区间，“RIFL”表示式（19）中的所提出的置信区间，“RIFL (80%)”表示采用 80%规则的所提出的 RIFL 置信区间。结果基于 500 次模拟重复。左侧面板中的虚线表示名义覆盖率 0.95；右侧面板中的虚线表示已知主导集合时预言机置信区间的宽度。

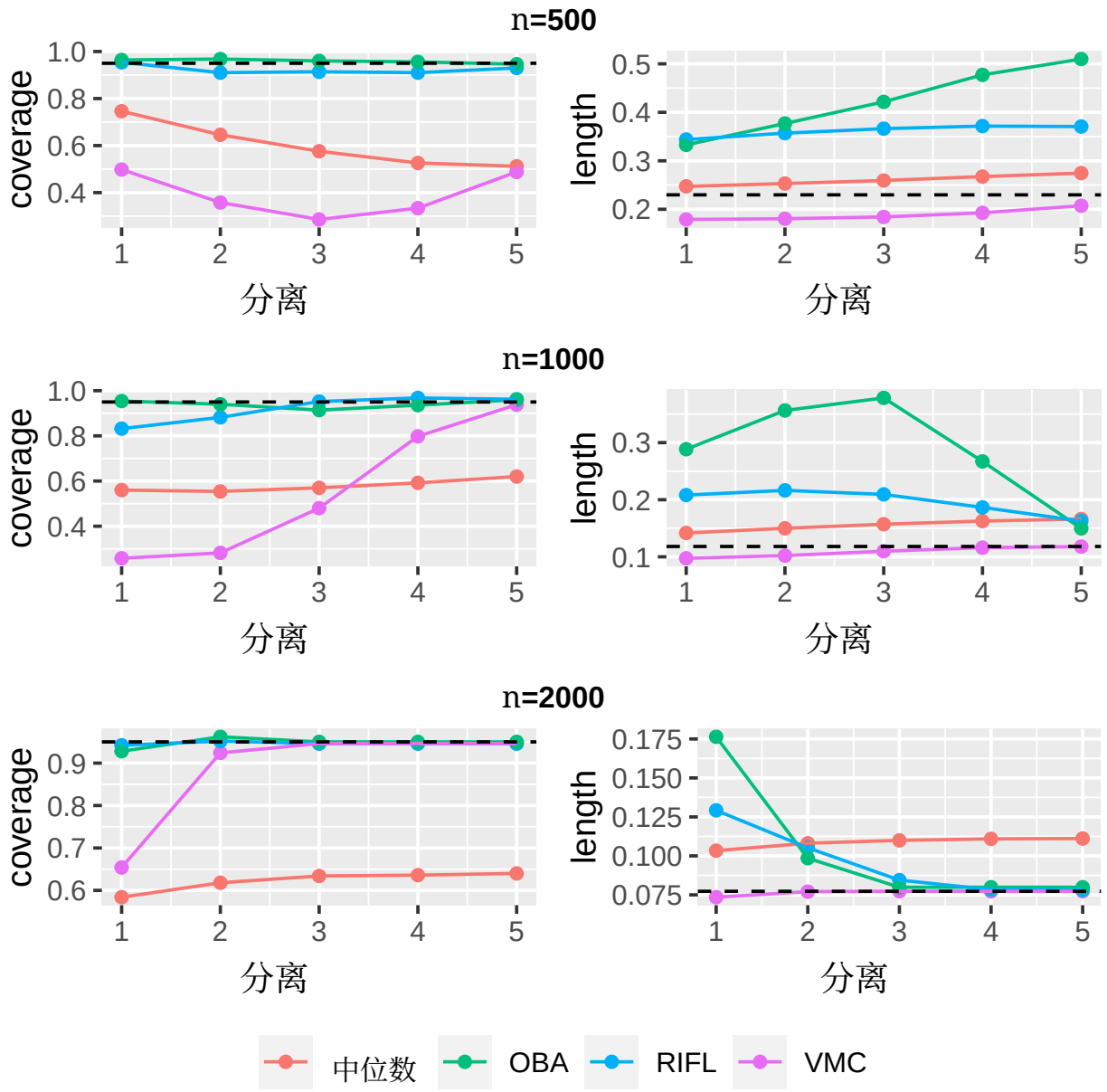

图 C3：高维预测：$\beta^* = \theta^*_8$ 的95%置信区间覆盖率和长度，包含 6 个多数站点，分离程度各异，其中 1 表示最低（最难检测），5 表示最高（最易检测）。“median”表示基于中位数估计量的置信区间，“VMC”表示最大团投票估计量及其在 (10) 中的相应置信区间，“OBA”表示 (39) 中的预言机偏差感知置信区间，“RIFL”表示我们在 (19) 中提出的置信区间，“RIFL (80%)”表示采用 80%规则的所提出的 RIFL 置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

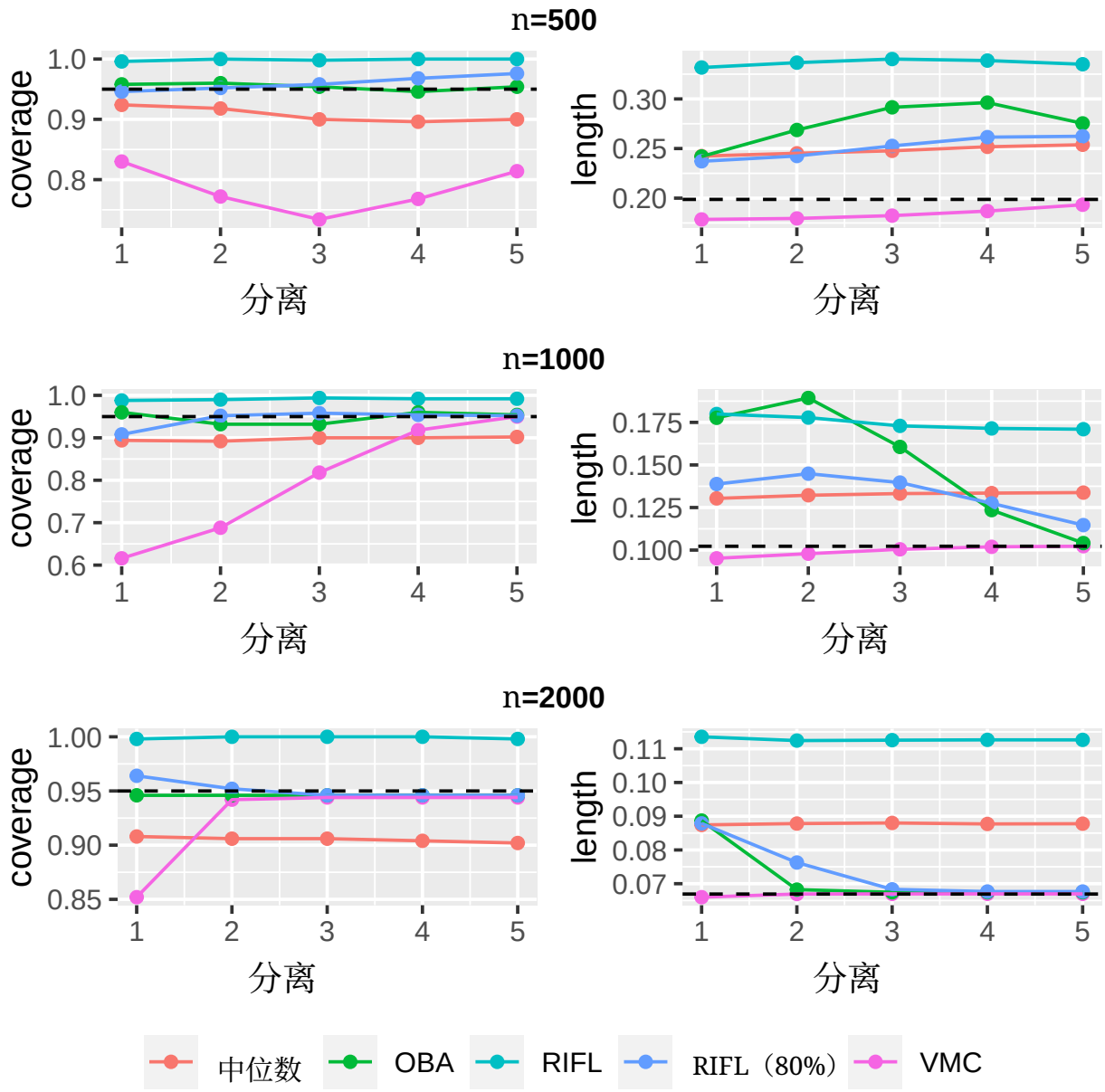

图 C4：高维预测：$\beta^* = \theta_8^*$ 在 8 个多数站点和不同分离程度下的 95%置信区间覆盖率和长度，其中 1 表示最低（最难检测），5 表示最高（最容易检测）。“median”表示基于中位数估计量的置信区间，“VMC”表示式 (10) 中的最大团投票估计量及其相应的置信区间，“OBA”表示式 (39) 中的预言机偏差感知置信区间，“RIFL”表示式 (19) 中我们提出的置信区间，“RIFL (80%)”表示利用 80%规则的 RIFL 置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

最后，我们给出第5.3节和第6.3节讨论的多源因果推断问题的结果，其中有8个多数站点。回想一下，第 l个站点的结局按照以下方式生成：

$$Y_i^{(l)} = \mu_l + [X_i^{(l)}]^\top \zeta^{(l)} + \beta^{(l)} A_i^{(l)} + \varepsilon_i^{(l)}, \quad \varepsilon_i^{(l)} \sim N(0,1).$$

对于有8个多数站点的情况，我们为 $l \in \{1,2,\ldots,8\}$、 $\beta^{(9)} = -1 - 0.2a$ 和$\beta^{(10)} = -1 - 0.1a$设置 $\beta^{(l)} = \beta^* = -1$，并让 a 在 $\{1,2,3,4,5\}$中变化。模拟设置的其他方面均与第6.3节相同。结果汇总于图C5中，我们再次观察到类似的模式。

C.2. m-out-of-n 自助法中 ν 的选择

本节展示采用不同 m的 m-out-of-n 自助法得到的置信区间的经验覆盖率和平均长度。具体而言，我们考察 $m = n^\nu$，并让 ν 在$\{0.6, 0.7, 0.8, 0.9, 1\}$中变化。我们聚焦于第6.1节中的低维预测示例，其中有6个多数站点，且所有10个站点的 $n_l = 1000$ 均相同。其他模拟设置均与第6.1节所述相同。

结果见图 C6。我们观察到 $\nu \in \{0.8, 0.9, 1\}$ 的覆盖率呈现相似模式。此外，在这 3 个取值中，选择 $\nu = 0.8$ 通常会得到最短的置信区间。对于 ν 的所有取值，当分离程度较低至中等时，MNB 置信区间的覆盖率低于名义水平。

C.3. 真实数据分析中 *16* 个站点的可推广性度量

图 C7 展示了我们对 COVID-19 真实数据分析中全部 16 个站点进行的完整可推广性度量。站点 4 的可推广性通常低于其他站点。当我们聚焦于一个特定风险因素（对应图 C7 中图的一行）时，有些站点的可推广性较低，这表明它们可能不属于主导集合。

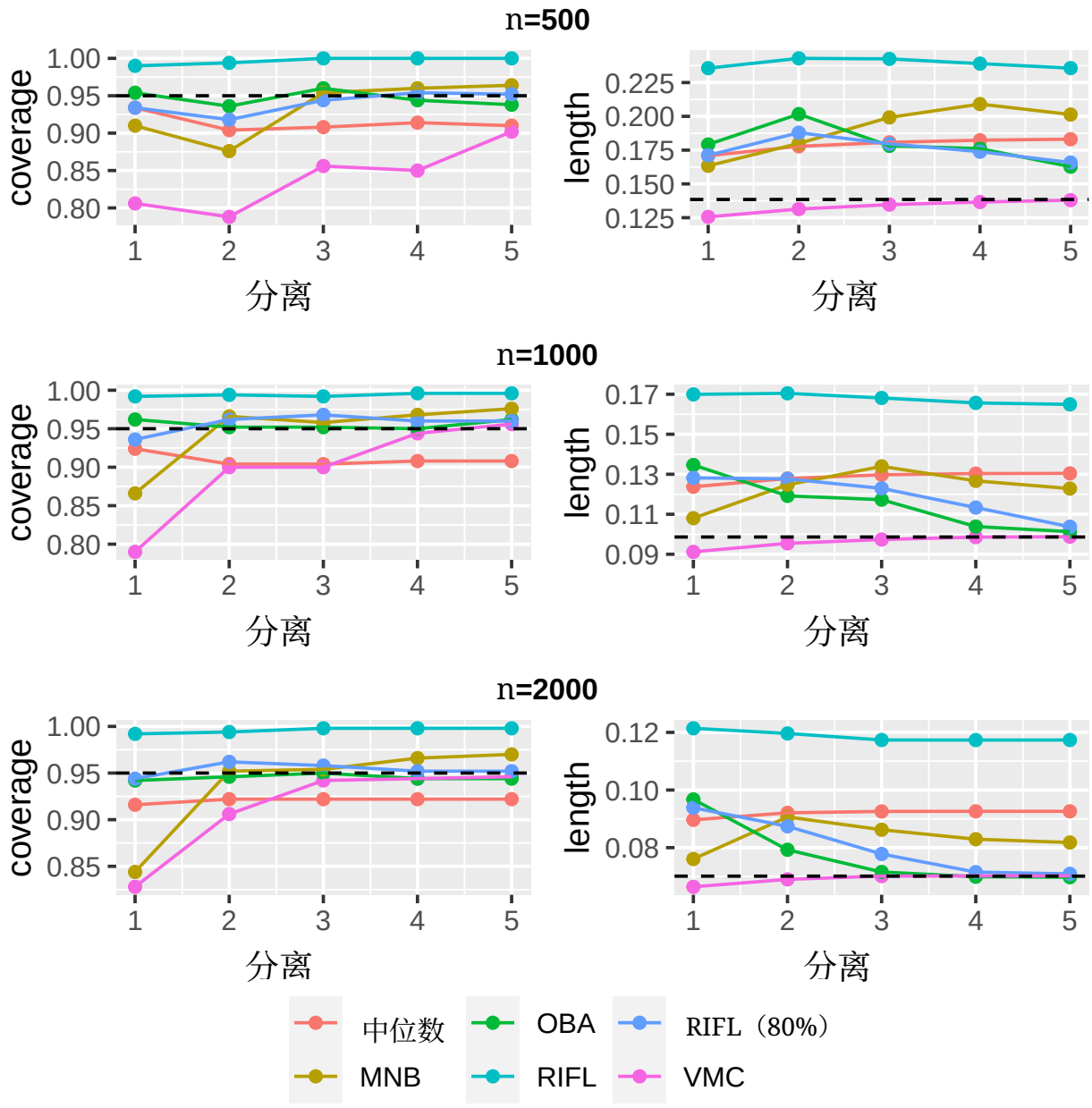

图 C5：因果推断：在有 8 个多数站点且分离程度各异的情况下，主导站点目标平均处理效应的 95%置信区间覆盖率和长度，其中 1 表示分离程度最低（最难检测），5 表示最高（最容易检测）。“median”表示基于中位数估计量的置信区间，“MNB”表示式（38）中的 m-out-of-n 自助法置信区间，“VMC”表示式（10）中的最大团投票估计量及其相应的置信区间，“OBA”表示式（39）中的预言机偏差感知置信区间，“RIFL”表示式（19）中我们提出的置信区间，“RIFL (80%)”表示利用 80%规则的 RIFL 置信区间。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95；右侧面板中的虚线对应于已知主导集合时预言机置信区间的宽度。

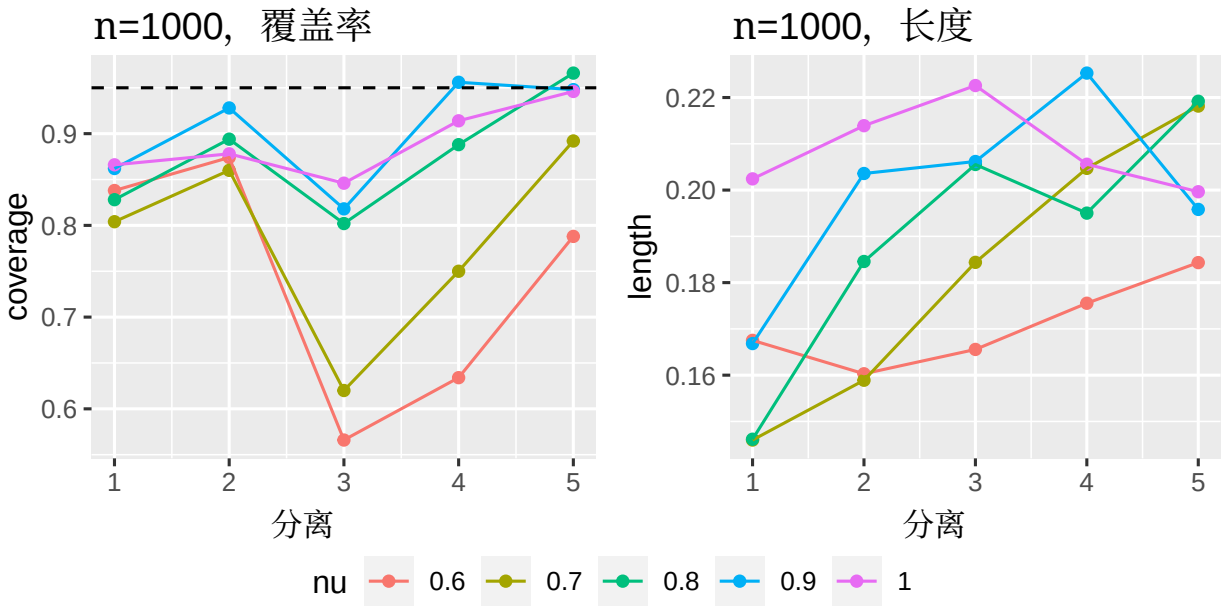

图 C6：低维预测：$\beta^* = \theta_1^*$ 的 m-out-of-n 自助法（MNB）置信区间（CI）在不同 m 选择下的覆盖率和长度。我们设定 $m = n^\nu$，并在 $\{0.6, 0.7, 0.8, 0.9, 1\}$ 中改变 ν。10 个站点的样本量均为 1,000，其中有 6 个多数站点。结果基于 500 次模拟重复。左侧面板中的虚线对应名义覆盖率 0.95。

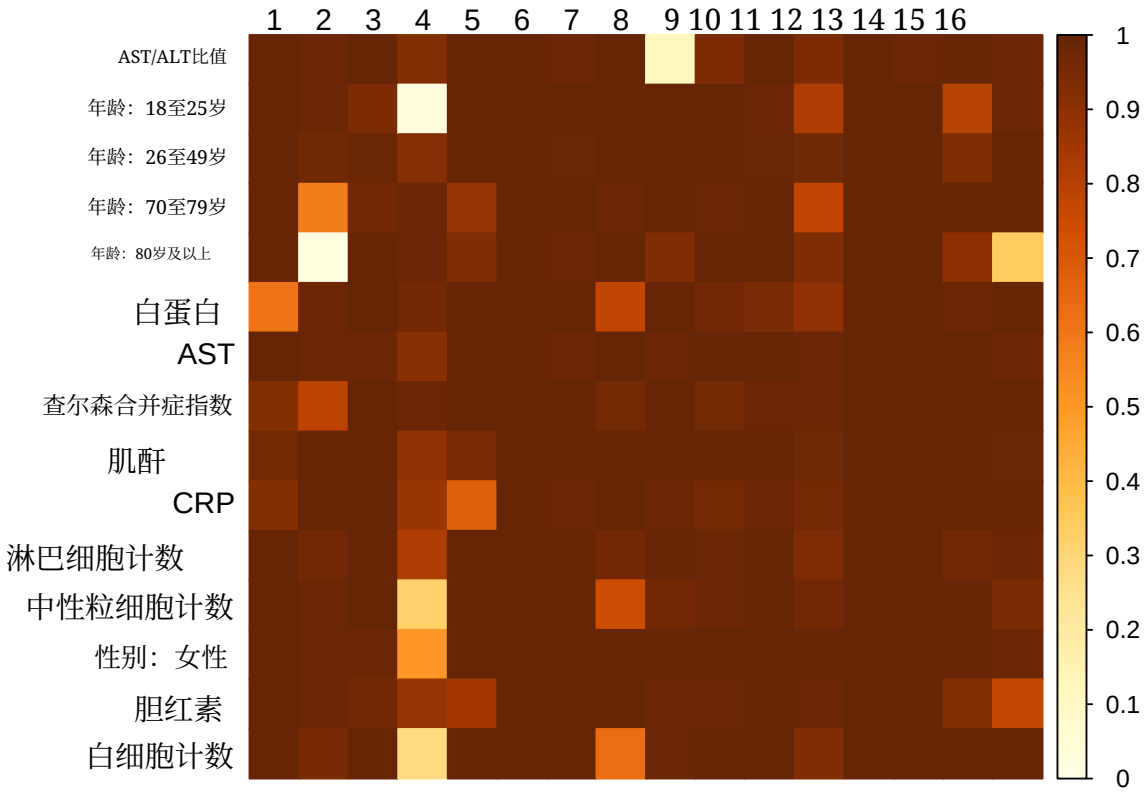

图 C7：15 个风险因素中每个因素在全部 16 个站点上的可推广性度量。每行对应一个风险因素，每列对应一个站点。